SUMMARY TECHNICAL REPORT
OF THE
NATIONAL DEFENSE RESEARCH COMMITTEE

[redacted]

Manuscript and illustrations for this volume were prepared for publication by the Summary Reports Group of the Columbia University Division of War Research under contract OEMsr-1131 with the Office of Scientific Research and Development. This volume was printed and bound by the Columbia University Press.

Distribution of the Summary Technical Report of NDRC has been made by the War and Navy Departments. Inquiries concerning the availability and distribution of the Summary Technical Report volumes and microfilmed and other reference material should be addressed to the War Department Library, Room 1A–522, The Pentagon, Washington 25, D. C., or to the Office of Naval Research, Navy Department, Attention: Reports and Documents Section, Washington 25, D. C.

Copy No.

68

This volume, like the seventy others of the Summary Technical Report of NDRC, has been written, edited, and printed under great pressure. Inevitably there are errors which have slipped past Division readers and proofreaders. There may be errors of fact not known at time of printing. The author has not been able to follow through his writing to the final page proof.

Please report errors to:

JOINT RESEARCH AND DEVELOPMENT BOARD
PROGRAMS DIVISION (STR ERRATA)
WASHINGTON 25, D. C.

A master errata sheet will be compiled from these reports and sent to recipients of the volume. Your help will make this book more useful to other readers and will be of great value in preparing any revisions.

[redacted]

SUMMARY TECHNICAL REPORT OF THE
APPLIED PSYCHOLOGY PANEL, NDRC

VOLUME 1

HUMAN FACTORS IN MILITARY EFFICIENCY

APTITUDE AND CLASSIFICATION

OFFICE OF SCIENTIFIC RESEARCH AND DEVELOPMENT
VANNEVAR BUSH, DIRECTOR

NATIONAL DEFENSE RESEARCH COMMITTEE
JAMES B. CONANT, CHAIRMAN

APPLIED PSYCHOLOGY PANEL
CHARLES W. BRAY, CHIEF

WASHINGTON, D. C., 1946

[redacted]

NATIONAL DEFENSE RESEARCH COMMITTEE

James B. Conant, *Chairman*

Richard C. Tolman, *Vice Chairman*

Roger Adams	Army Representative[1]
Frank B. Jewett	Navy Representative[2]
Karl T. Compton	Commissioner of Patents[3]

Irvin Stewart, *Executive Secretary*

[1]*Army representatives in order of service:*

Maj. Gen. G. V. Strong
Col. L. A. Denson
Maj. Gen. R. C. Moore
Col. P. R. Faymonville
Maj. Gen. C. C. Williams
Brig. Gen. E. A. Regnier
Brig. Gen. W. A. Wood, Jr.
Col. M. M. Irvine
Col. E. A. Routheau

[2]*Navy representatives in order of service:*

Rear Adm. H. G. Bowen
Rear Adm. J. A. Furer
Capt. Lybrand P. Smith
Rear Adm. A. H. Van Keuren
Commodore H. A. Schade

[3]*Commissioners of Patents in order of service:*

Conway P. Coe
Casper W. Ooms

NOTES ON THE ORGANIZATION OF NDRC

The duties of the National Defense Research Committee were (1) to recommend to the Director of OSRD suitable projects and research programs on the instrumentalities of warfare, together with contract facilities for carrying out these projects and programs, and (2) to administer the technical and scientific work of the contracts. More specifically, NDRC functioned by initiating research projects on requests from the Army or the Navy, or on requests from an allied government transmitted through the Liaison Office of OSRD, or on its own considered initiative as a result of the experience of its members. Proposals prepared by the Division, Panel, or Committee for research contracts for performance of the work involved in such projects were first reviewed by NDRC, and if approved, recommended to the Director of OSRD. Upon approval of a proposal by the Director, a contract permitting maximum flexibility of scientific effort was arranged. The business aspects of the contract, including such matters as materials, clearances, vouchers, patents, priorities, legal matters, and administration of patent matters were handled by the Executive Secretary of OSRD.

Originally NDRC administered its work through five divisions, each headed by one of the NDRC members. These were:

Division A—Armor and Ordnance
Division B—Bombs, Fuels, Gases, & Chemical Problems
Division C—Communication and Transportation
Division D—Detection, Controls, and Instruments
Division E—Patents and Inventions

In a reorganization in the fall of 1942, twenty-three administrative divisions, panels, or committees were created, each with a chief selected on the basis of his outstanding work in the particular field. The NDRC members then became a reviewing and advisory group to the Director of OSRD. The final organization was as follows:

Division 1—Ballistic Research
Division 2—Effects of Impact and Explosion
Division 3—Rocket Ordnance
Division 4—Ordnance Accessories
Division 5—New Missiles
Division 6—Sub-Surface Warfare
Division 7—Fire Control
Division 8—Explosives
Division 9—Chemistry
Division 10—Absorbents and Aerosols
Division 11—Chemical Engineering
Division 12—Transportation
Division 13—Electrical Communication
Division 14—Radar
Division 15—Radio Coordination
Division 16—Optics and Camouflage
Division 17—Physics
Division 18—War Metallurgy
Division 19—Miscellaneous
Applied Mathematics Panel
Applied Psychology Panel
Committee on Propagation
Tropical Deterioration Administrative Committee

NDRC FOREWORD

As events of the years preceding 1940 revealed more and more clearly the seriousness of the world situation, many scientists in this country came to realize the need of organizing scientific research for service in a national emergency. Recommendations which they made to the White House were given careful and sympathetic attention, and as a result the National Defense Research Committee [NDRC] was formed by Executive Order of the President in the summer of 1940. The members of NDRC, appointed by the President, were instructed to supplement the work of the Army and the Navy in the development of the instrumentalities of war. A year later, upon the establishment of the Office of Scientific Research and Development [OSRD], NDRC became one of its units.

The Summary Technical Report of NDRC is a conscientious effort on the part of NDRC to summarize and evaluate its work and to present it in a useful and permanent form. It comprises some seventy volumes broken into groups corresponding to the NDRC Divisions, Panels, and Committees.

The Summary Technical Report of each Division, Panel, or Committee is an integral survey of the work of that group. The first volume of each group's report contains a summary of the report, stating the problems presented and the philosophy of attacking them, and summarizing the results of the research, development, and training activities undertaken. Some volumes may be "state of the art" treatises covering subjects to which various research groups have contributed information. Others may contain descriptions of devices developed in the laboratories. A master index of all these divisional, panel, and committee reports which together constitute the Summary Technical Report of NDRC is contained in a separate volume, which also includes the index of a microfilm record of pertinent technical laboratory reports and reference material.

Some of the NDRC-sponsored researches which had been declassified by the end of 1945 were of sufficient popular interest that it was found desirable to report them in the form of monographs, such as the series on radar by Division 14 and the monograph on sampling inspection by the Applied Mathematics Panel. Since the material treated in them is not duplicated in the Summary Technical Report of NDRC, the monographs are an important part of the story of these aspects of NDRC research.

In contrast to the information on radar, which is of widespread interest and much of which is released to the public, the research on subsurface warfare is largely classified and is of general interest to a more restricted group. As a consequence, the report of Division 6 is found almost entirely in its Summary Technical Report, which runs to over twenty volumes. The extent of the work of a Division cannot therefore be judged solely by the number of volumes devoted to it in the Summary Technical Report of NDRC: account must be taken of the monographs and available reports published elsewhere.

The Applied Psychology Panel, under the direction first of W. S. Hunter and later of C. W. Bray, comprised a small group of psychologists and personnel specialists whose responsibility was to aid in refining and standardizing Army and Navy personnel procedures. The Panel devised selection and classification tests; it developed training methods; it improved the design of much equipment. The work of the Panel proved that it is as important to get the right man for a military job as it is to get the right ammunition for his gun.

The achievements of the Applied Psychology Panel cannot be measured in quantitative terms. But one can, for example, estimate with certainty that the tests devised to eliminate the emotionally unfit from induction prevented the wrecking of many lives and the fruitless expenditure of much time, effort, and money; and one can know surely that many lives were saved as the result of the one study alone which showed that the best night lookouts at sea were four times as proficient as the poorest.

The Summary Technical Report of the Panel, prepared under the direction of the Panel Chief and authorized by him for publication, is a record of scientific accomplishment and of zealous effort by able men working to increase the effectiveness of the nation's military manpower in time of national peril. The members of the Panel have our gratitude.

Vannevar Bush, Director
Office of Scientific Research and Development

J. B. Conant, Chairman
National Defense Research Committee

[redacted]

FOREWORD

What are the major problems which the Army and Navy must solve if sound strategy and tactics are to win a war? One is the problem of personnel. The Navy's Bureau of Naval Personnel and the General Staff evolved by years of military experience in the Army attest to the importance of personnel problems. The names of the General Staff divisions themselves indicate the extent of Service interest in the efficient use of manpower—G-1, Personnel; G-2, Intelligence; G-3, Training; and G-4, Supply. The interest in man is perennial. New methods of warfare, new explosives, new devices appear on the military scene, serve their period of usefulness, and become obsolete; the officer and the enlisted man remain.

During the war years 1940 to 1942 the Services turned to scientists for help on personnel problems just as they did for help on matériel problems. Psychological sections were organized in various branches of the Army and Navy. In 1940, the Adjutant General set up a Personnel Procedures Section. In 1941, the Air Surgeon created a Psychological Division. Early in 1942, the Navy's Bureau of Medicine and Surgery established an Aviation Psychology Section. Later that year the Bureau of Naval Personnel organized a personnel research program.

Within the National Defense Research Committee [NDRC] civilian scientists and engineers, engaged in research and development of military devices, also turned to the psychologist. Man is such an important factor in the successful use of military equipment that there was early recognition of the need to study the man-machine combination, not the machine alone and not the man alone. Research on fire control, underwater sound, and communications equipment included psychological research on the human operators of these types of equipment. In night operations, too, the significance of the human factor was evident, and NDRC undertook psychological research on night vision.

These scattered activities in and out of the Services proved the value of psychological research in meeting some Army and Navy problems. Measurable increases in the efficiency of personnel followed the application of psychological techniques. These improvements were of interest to nearly all branches of the Services and to a number of divisions of NDRC. The result was that psychological research units came to be sponsored by a remarkable variety of organizations: administrative, medical, engineering, physical, and mathematical.

In spite of the diversity of its sponsorship, psychological research was always psychological, studying the man in relation to the machines he had to operate. Experience proved that the same knowledge and the same experimental methods helped to solve military problems which at first appeared quite unlike each other. Thus the merits of a single coordinating psychological research unit gradually became apparent. There was need for such a group to encourage the development of an overall, professional approach to military psychology. There was need for such a group, not to take over the many scattered groups already in existence, but to supplement their work by entering into fields which their limited directives and cognizances prevented them from entering.

In direct recognition of this need the Navy Department, with the War Department concurring, on June 4, 1942 requested NDRC to establish the Committee on Service Personnel—Selection and Training. This committee was formed in the National Research Council under an NDRC contract. The civilian members of the Committee consisted of John M. Stalnaker, Chairman, George K. Bennett, Leonard Carmichael, Clarence H. Graham, and Morris S. Viteles. Walter V. Bingham and P. E. McDowell served as Army and Navy representatives. The writer was the committee's executive secretary. He was assisted by John L. Kennedy.

In October 1943, the Committee on Service Personnel became the Applied Psychology Panel of NDRC. The Panel membership consisted of the civilian members of the predecessor committee with the addition of Walter S. Hunter, who served as Chief of the Panel until the Japanese surrender, and Dael Wolfle. John L.

Kennedy and the writer were the technical aides. Later, Dael Wolfle also became a technical aide.

The first work of the Committee was almost entirely directed toward the solution of problems which had high priority because of the shortage of manpower or the importance of the job itself, but which were not being studied by preexisting groups of psychologists. These spot programs concerned the selection, classification, and training of personnel. As the work progressed, emphasis gradually shifted to attacks on more fundamental problems of simplifying selection, classification, and training procedures through improvements in the design and operation of equipment. In the course of this shift many of the psychological projects came to be liaison as well as research centers, in which engineers, scientists, personnel men, and combat officers brought together their problems of man-machine efficiency and developed coordinated research programs on many aspects of military psychology.

The experimental development and evaluation of personnel procedures led to marked improvement in Service efficiency. The two volumes of this Summary Technical Report contain the evidence that scientific methods can solve personnel problems in gunnery, fire control, communications, radar, and night operations; in surface, subsurface, ground, aerial, and amphibious units; in elementary, basic, advanced, and operational training.

The reader who studies the detailed record contained in the Panel's Summary Technical Report will note one characteristic which is common to all successful research in military psychology. Studies of classification, training, and equipment, diverse though these subjects may appear to be, all rely on a common tool, the measurement of human performance on the job. The psychologist calls such a measurement the "criterion." The criterion is the measuring stick with which he evaluates the success of a proposed test or procedure. Psychological tests, training aids and procedures, and designs for equipment, however sound they may be from a theoretical point of view, are valuable only in proportion as they improve human performance. That performance must be *measured* if one is to know whether or not improvement has actually occurred. Three examples of the use of the criterion in military psychology are given in order to define the nature of psychological science today.

When the Army and Navy asked the Panel to construct a test for the selection of radio code operators, the first problem was to develop a sensible, reliable, accurate, and objective criterion measure of men's ability to receive code. Without such a criterion measure, there was no basis for knowing which of the already existing code aptitude tests was best or of knowing whether new tests were any better than the old ones. The development of a good criterion made it possible to construct a new code aptitude test, better than any previous ones. The new test was adopted by both Army and Navy.

When the Navy asked the Panel to evaluate several procedures for transmitting code messages over radar systems, the same kind of questions had to be answered. In this case, however, the problem was to define the best rate at which to send messages and the best way of training radar operators to receive messages on their oscilloscopes. A solution based on fact rather than on opinion was possible only by measuring the proficiency of various groups of radar operators in receiving code at various rates after various kinds of training.

When the Army Air Forces requested the Panel to improve the controls of the B-29 gunsight, the question of a criterion arose again. An endless variety of controls was possible. Each suggestion had its advocates, but because the accuracy of aerial gunnery was so difficult to measure, choice among the alternatives was based on opinion instead of fact.

The controls of a standard B-29 gunsight were attached to writing pens on a strip of paper. The paper moved at a constant speed under the pens. When the gunsight moved, the pens moved proportionally. With this device the gunner tracked a synthetic target. The ink record furnished a measure of the proficiency of the gunner.

With this new criterion, one defective control was spotted at once. Each time that a gunner pressed his trigger the gunsight jerked 10 or

even 20 mils to one side or the other. In the remote control gunnery system of the B-29, the effect of pressing the trigger was not due to recoil of the guns. It was due to the relation between the gunner and his gun controls. The controls were improperly arranged for good triggering performance.

The record of gunner performance permitted the evaluation of new controls. Choice of the best controls in terms of their relation to gunner performance was possible. The new record also permitted the Panel to recommend a new operating procedure for gunners using the standard sight: Fire continuously when once you have begun to fire. It stimulated the development of an automatic burst-control trigger which fired the guns intermittently even though the gunner pressed the trigger continuously. The record permitted the evaluation of alternative methods of training gunners. The same measure is available as a criterion for evaluating aptitude tests for gunners.

Throughout this Summary Technical Report the problem of securing an objective and reliable measure of performance on the job recurs again and again. Whether the problem is to classify men, to train them, or to adjust operating conditions for maximal efficiency, success depends upon the ability to measure accurately what men do.

Considerable ingenuity was shown during World War II in developing measures of proficiency, but too many of those measures were devised for school and training use. More effective work could have been done if more opportunities had been created for studying combat performance, the ultimate goal of all military research. In fact, the only phase of military operations which has so far not been directly studied on a large scale by the psychologist is combat itself. Better selection, more realistic training, and more efficient operating procedures will become available when combat performance is measured and used as a criterion.

In the years to come the Army and Navy may have practically any kind of research program in the field of military psychology that they desire. Psychological work on military personnel during World War II varied from none at all to a thorough research analysis of the whole interrelated set of problems involved in selecting men for special duty, training those men for the duties they were to perform, and modifying the equipment in order to make it easier to operate and easier to learn to operate. The various possibilities for the future are enumerated below.

Prior to World War II it was the frequent experience of psychologists interested in military problems to be told that the Army and Navy welcomed the opportunity to consult with them but needed no research assistance. This attitude persisted throughout the war years in the minds of some officers but was by no means universal, for research was not only welcomed but eagerly sought in a number of commands and on a number of problems.

Both Army and Navy do know a great deal about personnel. Both have a long history of experience and a record of success. In addition, both have the benefit of recent advice from expert consultants and the results of wartime research studies to aid in handling personnel during peacetime. It might therefore be decided, either as a matter of general policy or for a particular field, that no further research is necessary.

The Applied Psychology Panel on a number of occasions was asked to furnish a man well acquainted with a particular type of military problem who could serve as an expert consultant for a few days' time. During these short periods of consultation the man was expected to survey, for instance, a training installation, observing the instruction and recommending improvements in instructional material, course organization, lesson plans, and examinations.

The results of such consultations varied, depending upon the experience of the consultant and upon the similarity between the problems involved in the military situation and those with which he had had previous experience.

One successful instance was the Applied Psychology Panel's work on improving the training of Combat Information Center [CIC] personnel under Commander Operational Training Command, United States Pacific Fleet [COTCPac]. In this case it was possible to provide COTCPac with a consultant who had wide and varied experience with radar equipment, with radar operator training in the Services, and with educa-

tional methods. In a few days' time spent at each CIC school, this consultant was able to eliminate a number of difficulties and to suggest improvements which were eagerly received by the instructors in charge.

The usefulness of consultation depends on the degree of familiarity of the advisor with actual Service conditions. The best of men may come into a Service situation with no direct experience, or experience only from the remote past, of the details of Service operations. He can deal only in generalities. These do not compare in value with the specific, detailed recommendations which are possible to an expert who has had the benefit of months of research work in a training or operating area.

Consultation on special problems is the second type of opportunity available to the Army and Navy.

As a third possibility, psychologists can be used to maintain already established personnel objectives and procedures. Psychological procedures, like others, require continuing attention or they gradually become ineffective. Failure to maintain personnel methods results in a loss of the values originally attained.

A program of maintenance is necessary. If, however, research is limited to maintenance alone, any mistake in original plans may persist unchallenged and new developments may pass unheeded.

The final type of study which may be conducted consists of a thoroughgoing continuing research analysis of three interrelated sets of problems. These problems are the ones involved in the classification of personnel for particular duties, the problems involved in their training, and the problems involved in designing the equipment which they use in order to make both learning and operation easier and more efficient.[a]

Working simultaneously on selection, training, and equipment problems is the most efficient type of program and may normally be expected to lead to the greatest improvements. As shown above, the research methods are the same for all three. And the three kinds of problems are all centered in the human operator; they are necessarily interrelated. Changing selection standards may permit a shorter, or require a longer, training period. Simplifying the operation of a piece of equipment may allow a relaxation of selection standards and require a modification of training plans. Working simultaneously on selection, training, and equipment allows for the efficient coordination of plans concerning all three problems.

It seems clear that the most effective research for the future Army and Navy will be research on problems related to new equipment. Research on the classification and training of personnel and on methods of operation of equipment will be most effective if results are available when new devices roll off the production line. If studies of equipment design are completed before production begins it will be unnecessary to suggest field modifications or second models. The best time for psychological research is before the need for it is obvious. It is only necessary to add one or a few additional mock-ups and preproduction models to the five or six now commonly constructed to make psychological research possible in the preproduction period.

The direction of psychological research toward new equipment suggests that psychological research should be associated with other research and development work. Only by associating psychological research with all research will there be automatic provision that personnel requirements will be subjected to intensive study at the most effective time.

The Army and the Navy can have any type of assistance they desire. If they desire a complete program they must secure men with adequate training in experimental psychology. Whether psychologists will accept Service positions depends upon the conditions offered for experimental work.

Psychologists will not enter the Army or Navy during peacetime unless there is good assurance that they will be able to work effectively on psychological problems. That assurance depends upon freedom from the danger of being

[a] Since the Panel worked only on classification, training, and equipment, no other aspects of military psychology are considered here. If a complete research program is contemplated, consideration should also be given to the inclusion of research on the abnormal individual and on the social psychology of morale, and psychological warfare.

sidetracked into routine administrative duty. It depends upon long-term planning; year-by-year commitments are acceptable during wartime, but not in planning peacetime research. It depends upon the ability of the military psychologist to maintain his status as a recognized member of the profession of psychology. It depends upon the creation of a single psychological command for psychological research within a Service or major branch of a Service. It depends upon a fairly high staff level of military supervision so that psychological planning and direction will be in the hands of officers with an appreciation of the broad implications of psychological research and with the authority to see that suitable research facilities are made available.

Unless the Army and Navy are able to offer assurance that psychologists will be able to work effectively on psychological problems, the several hundred positions now being created for military psychologists will remain unfilled. At the time of writing this foreword only a handful of the thousand or more psychologists engaged in wartime psychological work have accepted permanent positions with the Services. Salaries, frequently 50 per cent higher than the universities offer, are not attractive unless they are accompanied by opportunities to work effectively and continuously on important psychological problems, opportunities to see the results incorporated into Service practice, and opportunities to maintain close relations with civilian colleagues.

The Army and Navy can have any type of assistance they desire. Whether either Service will have an effective research program in military psychology depends upon the availability of personnel with adequate training in experimental psychology. Competent psychologists will be available when satisfactory conditions for experimental work are known to exist.

Service interest in man is perennial. Service recognition of this fact is illustrated by the words of a battle-scarred and ribbon-covered Admiral appointed to demonstrate the mechanical and electronic marvels of the USS *Missouri* to a group of distinguished scientists from NDRC. His opening words, as he swept his hand over the deck of the pride of the fleet, were these: "Twenty-five hundred officers and men: gentlemen, twenty-five hundred sources of error."

It was the purpose of the Applied Psychology Panel to help the Army and Navy reduce human error during World War II. The success of future efforts to reduce human error—the success of the future development of military psychology—rests with the Army and the Navy.

CHARLES W. BRAY
Chief, Applied Psychology Panel

[redacted]

PREFACE

THE SUMMARY TECHNICAL REPORT is a systematic account of the work done under the direction of the Applied Psychology Panel. Volume I describes selection and classification of military personnel; Volume II describes military training and the human factors involved in the design and operation of military equipment. In each of these three fields—selection and classification, training, and the design and operation of military equipment—the work actually done and the effects of that work on military practice are described.

In the foreword to this volume Dr. Charles W. Bray, Chief of the Applied Psychology Panel, has described the Panel's recommendations for future research in military psychology.

This volume and the Panel's recommendations for future work on classification are summarized in Chapter 1. Chapters 2 to 13 give details of the Panel's work on the selection and classification of military personnel. These chapters may be read in any order. Each gives an account of one phase of the work. Chapters 14 and 15 will give the professional psychologist an understanding of the psychological principles involved and the methods followed in the development and evaluation of personnel classification devices.

In Volume II, Chapter 1 gives a brief factual summary of the Panel's work on military training and on the development of design improvements and standard operating procedures for several types of military equipment. Chapters 2 to 12 describe specific training studies. Chapters 13 to 17 cover general principles of military training. Chapters 18 to 23 discuss the design and operation of special types of military equipment. The two final chapters, 24 and 25, present some guiding principles for future equipment development.

Cross references are given by means of section numbers, for example 11.3.5, in which the 11 refers to the chapter, 3 to the third major division of Chapter 11, and 5 to the fifth section of the third division of Chapter 11. Commonly used abbreviations are explained in a glossary at the end of the volume.

In writing this final summary a few tables have been recalculated from the original reports. In no case were the changes large enough to alter conclusions or recommendations.

The Applied Psychology Panel has had help from many sources in preparing this final account of its work. The Army, the Navy, the Applied Psychology Panel contractors, and the Bausch and Lomb Company have provided photographs to illustrate many of the devices and procedures discussed. To each of these we express our thanks.

The author of each chapter is named in the table of contents and at the beginning of the chapter. Though all chapters are based upon the original reports prepared by the contractors, many of them were finally written by the editor. Approximately half of the chapters were written by men who spent the war years in the field, working on problems their chapters summarize. The Panel and the editor are very appreciative of the time and effort devoted by these men to writing firsthand accounts of their work.

Credit is not given in the Summary Technical Report to the individual psychologists who actually made the contributions here reported. Their specific contributions can be discovered only by studying the original reports listed in the bibliographies. The Applied Psychology Panel expresses its sincere appreciation to these men for their individual contributions and for their effective teamwork. These two volumes record their achievement.

DAEL WOLFLE
Editor

CONTENTS

Chapter 1

SUMMARY

By *Dael Wolfle*

1.1 INTRODUCTION

THIS CHAPTER is a summary of the work of the Applied Psychology Panel, NDRC, on problems of military selection and classification—the work reported in the remaining chapters of this volume. The chapter is organized into four sections. The first outlines the Panel's contributions to the development of tests for use in selecting military specialists. The second describes the Navy classification program and the Panel's contribution to it. The third describes the methods employed in the construction of tests for the selection and classification of military personnel. The fourth brings together the Panel's recommendations for future development of military classification processes.

The subtitles of the following sections include the numbers of the chapters summarized in each.

1.2 TESTS FOR SELECTING MILITARY PERSONNEL

1.2.1 The United States Navy Basic Classification Test Battery (Chapter 2)

The U. S. Navy Basic Classification Test Battery consists of five tests. Project N-106 of the Applied Psychology Panel and the Test and Research Section of the Bureau of Naval Personnel cooperated in their development. The Test and Research Section had major responsibility for developing two of the tests, the Mechanical Aptitude Test (MAT) and the Mechanical Knowledge Test. The project assumed major responsibility for the initial development of the other three, the General Classification Test (GCT), the Reading Test (R), and the Arithmetical Reasoning Test (AR). The project developed two forms of each of these tests, produced an additional form of the GCT suitable for administration in the fleet, and prepared an additional test named the Arithmetical Computation Test (AC) for possible use in future general selection batteries.

Statistical information regarding the GCT, R, AR, and AC tests (and for all tests developed by the Panel) is given in Table 1. The first three, as part of the Basic Classification Battery, were used from the time of their completion at the end of 1942 until the end of World War II, in the initial classification and assignment of all Navy recruits.

1.2.2 The Officer Qualification Test (Chapter 3)

The Bureau of Naval Personnel requested a test which could be used to select those officer applicants who had the ability to complete Navy officer training courses satisfactorily. The request was met by developing three forms of the Officer Qualification Test (OQT). The OQT was used in the Offices of Naval Officer Procurement until officer recruitment ended at the close of World War II.

1.2.3 Predicting Combat Leadership (Chapter 3)

In fulfillment of a request from the Office of the Adjutant General, War Department, the Applied Psychology Panel conducted a preliminary study of the prediction of combat leadership. Two classes of infantry officer candidates and two classes of field artillery officer candidates were studied. In addition, combat ratings of 176 infantry company officers were compared with several types of officer candidate school (OCS) records.

Within the already-selected group of students in infantry OCS, it was found that none of the items of information obtainable before or upon entrance to the school was highly predic-

TABLE 1. Selection tests developed by projects of the Applied Psychology Panel.

Test	No. of forms	Purpose	Reliability	Validity	Service use	Details in chapter
General Classification Test (GCT)	2	For use in general classification of naval recruits to decide on type of duty or special training	.94	.32–.63* .27–.63	Routine Navy use (NavPers 16502)	2
Reading (R)	2		.85	.32–.59* .26–.46	Routine Navy use (NavPers 16512)	2
Arithmetical Reasoning (AR)	2		.83	.28–.63* .27–.51	Routine Navy use (NavPers 16512)	2
Arithmetical Computation (AC)	1	Experimental test for possible inclusion in Basic Battery	...	.33–.69†	Experimental use only; not adopted	2
Officer Qualification Test (OQT)	3	To predict ability to complete Navy officer training courses satisfactorily	.90	.48–.51‡ .33–.47	Routine Navy use (NavPers 16561, 16563)	3
Personal Inventory (PI)	3	To identify emotionally unstable men likely to be discharged for psychiatric cause	.66–.92	See Chapter 4	Routine use by Navy (NavPers 16845), U.S. Coast Guard, U. S. Maritime Service. Adapted by AAF	4
Vocational Interest Inventory	1	To determine interests which might be used as an aid in military classification	.77–.94§	Undetermined	None	5
Radio Code Aptitude Speed of Response (SOR)	1	To select men for training as radio code operators	...	.50	Adopted for routine use by Army as *Army Radio Code Aptitude Test*, 1944, *ARC-1*, and by Navy as Radio Code Test: Speed of Response, Form 2.	6
Experimental forms of 13 tests designed to select radar operators			.81–.98	Two of the 13 included in Navy CIC Aptitude Test, Form 2 (NavPers 16980) which had validity of .45–.46		7
Speech Interview	Unlimited	To rate ability of men for duty as shipboard telephone talkers	...	Undetermined	Adopted by Navy for routine use	10

*There is, of course, no single validity for a test. The upper pair of values given are the highest and the lowest correlations found between scores on Form 1 tests and grades in 13 elementary Service schools. The lower pair are the highest and lowest correlations between scores on Form 2 tests and grades in 12 elementary Service schools. The correlations have been corrected for curtailment in range of talent.

†Range of correlations with grades in 12 elementary Service schools.

‡Upper figures are correlations with grades in Naval Training School (Indoctrination); lower figures are correlations with grades made by women in Naval Reserve Midshipman's School.

§Reliabilities of the five scores, each representing interest in a different area of behavior, varied from .77 to .94.

[redacted]

tive of later performance. Scores on the Army General Classification Test and leadership qualities as indicated by former civilian positions were the two most predictive variables.

Background information available upon entrance to field artillery OCS was generally more useful in predicting success in school than was true of infantry OCS. Previous education and grades on mathematics examinations showed useful relationships (correlations of .33 to .48) with success in field artillery OCS.

There was fair agreement between combat ratings and leadership ratings assigned while in infantry OCS. Of the other variables studied, only age had predictive value for combat success. The superior and excellent officers on the whole came from the age group 22 to 28. Proportionately fewer men below 22 or above 28 were rated superior or excellent by their regimental commanders, executive officers, or battalion commanders.

1.2.4 Eliminating the Emotionally Unfit (Chapter 4)

The Personal Inventory (PI) was developed in order to have available a device for making quick identification of those emotionally unstable men who were likely to break down under the stress of hazardous duty.

Three forms of the PI were developed.

1. A long form containing 145 items in each of which the man taking the test was forced to choose between two short descriptive statements, selecting the one which he felt more appropriately applied to him.

2. A short form consisting of 20 items selected from the long form. The 20 selected were those which experience had shown to be most useful in discriminating between acceptable men and men later discharged for psychiatric reasons. The correlation between the long and short forms was .84.

3. A long form suitable for use with officers.

The PI was used as a preliminary screening device to identify the men most likely to be given subsequent psychiatric discharges. When the flow of men was too great to permit careful psychiatric interviews of all men, the PI identified those who had to be interviewed most carefully. Validity of the PI was determined by measuring the success with which it could select in advance those men whom the psychiatrists later decided unfit for duty. Its usefulness as a screen consisted of reducing the number of men who had to be interviewed in order to find a given proportion of the men who would later be discharged. The number to be interviewed could vary depending upon the amount of interview time available. For example, the PI could select a group of 16 per cent of the total and in that 16 per cent include 60 per cent of those given discharges. If time permitted interviewing more men, the PI could select a group consisting of 42 per cent of the total and in the 42 per cent include over 75 per cent of those given discharges.

Preliminary evidence indicated that the PI could contribute to improved selection of amphibious forces officers, paratroopers, and aviation personnel. Its value in selecting Marine Corps officer candidates was slight.

The short form of the PI was used by the Navy as a preliminary psychiatric screen. The long form was used by the Coast Guard and the U. S. Maritime Commission. The Army Air Forces developed a special adaptation for their use.

1.2.5 Determining Vocational Interests (Chapter 5)

At the request of the Adjutant General, War Department, an inventory of interests was constructed for possible use in the classification of military personnel. The inventory was intended to secure information on a wide variety of interests, for example, vocational, social, recreational, scientific, and religious interests. Information of this type could provide a valuable supplement to that obtained by interviews, vocational history, or standardized tests of ability in classifying men for military duty.

The interest inventory, a scoring key, and detailed statistical information were turned over to the Adjutant General's Office at their request, for determination of its actual usefulness in military situations. No report has been received from that office.

[redacted]

1.2.6 Selecting Radio Code Operators (Chapter 6)

At Navy request the Applied Psychology Panel developed an improved test of radio code aptitude. The test first required the subject to learn three simple characters of International Morse Code and then measured his ability to differentiate those characters as they were sent at faster and faster speeds. Instructions, the learning part of the test, and the test itself were all recorded on phonograph records to ensure standardized administration. The test was officially adopted by both Army and Navy for use in the selection of men for radio code training.

1.2.7 Selecting Radar Operators (Chapter 7)

A number of tests were designed for selecting and screening radar operators. The tests were intended to measure aptitude for the visual tasks of reading and interpreting oscilloscope patterns. Two of the tests were adopted after revision by the Navy for use as two of the three parts of the Combat Information Center Aptitude Test, Form 2 (NavPers 16980). The third part was developed by the University of California Division of War Research under NDRC Section 6.1. The total test showed correlations of .45, .55, and .56 with final grades in three successive classes at Naval training school (tactical radar).

1.2.8 Selecting Stereoscopic Rangefinder and Heightfinder Operators (Chapter 8)

Standards for the selection of stereoscopic heightfinder operators were developed for, recommended to, and adopted by the Army. A Panel project demonstrated the validity of these standards. Assistance was given to the Army in setting up stereoscopic testing centers where men to be trained as heightfinder operators were selected.

When the Navy established a school for training men for the rate of fire controlmen (R), the Army heightfinder standards were used as the basis for setting selection requirements for rangefinder operators. The Applied Psychology Panel work was extended at that time to include studies of the selection and training of rangefinder operators. Improved selection standards were recommended to the Navy and adopted for general use. The Project staff trained Navy personnel for duty in the two testing centers where prospective rangefinder operators were selected.

In the course of developing selection standards for stereoscopic observers, information was obtained on the reliability, validity, and intercorrelations of a number of visual tests. Intercorrelations among the tests purportedly measuring stereoscopic ability were shown to be conspicuously lower than the test reliabilities. This finding suggests that the tests do not measure exactly the same visual function and that each test must measure at least some elements which are specific to itself alone. A thorough study of military visual requirements and of improved methods of visual testing was recommended.

1.2.9 Selecting Night Lookouts (Chapter 9)

Studies on the prediction of night lookout performance showed that the tests of night vision being used had low reliability and very low validity. One study, for example, showed no relation between scores on the radium plaque adaptometer and measured performance of the men as night lookouts. The Bureau of Naval Personnel directed thereafter that scores on the radium plaque adaptometer should not be recorded on the Enlisted Personnel Qualifications Card.

1.2.10 Selecting Navy Telephone Talkers (Chapter 10)

A speech interview to rate the potential ability of men as shipboard telephone talkers was developed. The interview was conducted over Navy sound-powered phones. It included the pronunciation of numbers and of all common

American speech sounds, repetition of commands, and extemporaneous description.

A training course was developed to teach classification petty officers to administer the speech interview and to rate men as "well qualified," "qualified," or "not qualified" for duty as telephone talkers.

The speech interview was adopted by the Bureau of Naval Personnel for routine use in the classification of enlisted personnel.

1.2.11 Summary of Tests Developed under the Applied Psychology Panel

Table 1 is a summary of the tests developed by projects of the Applied Psychology Panel for use in military selection. It gives pertinent statistical information and indicates the use made by the Army or Navy of each test.

1.3 THE CLASSIFICATION OF MILITARY PERSONNEL

The tests described in Chapters 2 to 10 were developed to aid in the selection of men who would be successful in particular military specialties or to aid in eliminating those unfit for special types of military duty. If the Army or Navy had an unlimited pool of men from which to select its specialists, each of these tests could select a group of men most of whom would satisfactorily learn their new duties. With a limited manpower supply, the military task is more complicated than that: all available men must be allocated. This process requires using every man and attempting to assign each to the type of duty in which he will be most successful.

1.3.1 The Objectives of Classification (Chapter 11)

The objectives of classification are to select able men for special training, to assign each man to the job he can do best, and to divide available manpower equitably among the various specialties and among the various units of a military force. The Applied Psychology Panel cooperated with the Bureau of Naval Personnel in three programs designed to develop procedures for achieving these objectives. First was the classification of the crew of USS *New Jersey.* Following this there were larger projects to classify men for the Amphibious Training Command of the U. S. Atlantic Fleet and for the crews of destroyers being manned on the West Coast. On the basis of the success of the three programs, the Commander-in-Chief, United States Fleet, directed the Chief of the Bureau of Naval Personnel to develop similar classification programs for fleet-wide application.

1.3.2 Devices to Aid Military Classification (Chapter 11)

Four types of devices were developed to aid in the process of military classification. One was the series of tests described above. The second was a series of special devices to aid the interviewer in securing relevant information about a man's experience and ability. The third was a mechanical system used on shipboard for sorting personnel record cards in order to locate quickly the men who possessed needed skills or traits. The fourth was a device, called the selectometer, which automatically weights and combines a man's scores (or other measures) on each of the factors considered important in selecting men for a particular job. The device differs from a human classification interviewer in being more consistent in the importance it attaches to each variable, in not forgetting any of the important traits, and in being able to weight and combine scores used in predicting a man's potential success in a number of different jobs simultaneously. Several models of the selectometer were built for use in Navy classification centers.

1.3.3 Organization of a Classification Program for Recruits (Chapter 12)

The procedures followed during World War II in classifying and assigning Navy recruits are briefly described. Five improvements in the classification procedures are recommended.

1. Using a primary test battery to separate the men of school quality from those who should be assigned to general detail and a secondary battery to determine the particular school to

which each man in the first group should be assigned.

2. Improving the interview by standardizing the weighting assigned to various factors on which the interviewer's recommendation should be based.

3. Improving the basis for filling school quotas by more general use of assignment pools and by better determination of proper priorities in filling quotas.

4. Continuous research to keep classification tests and methods abreast of changing military requirements.

5. Indoctrinating regular officers in modern methods of personnel classification.

1.3.4 Organization of an Advanced Classification Program (Chapter 13)

The methods used by the Navy in advanced classification and reclassification of enlisted personnel are briefly described. Five recommendations for improving advanced classification are offered.

1. More extensive construction and use of objective methods of determining the actual proficiency with which a man can perform the duties of a particular billet.

2. Improvement in the methods of rating shipboard proficiency.

3. Use of brief oral tests, called work readiness tests, to determine how much a man knows about the details of a particular type of duty.

4. The indoctrination of regular officers in the methods and information available for use in making shipboard assignments.

5. Improved methods for coding and filing personnel data to make those data more easily available and more likely to be used in ship and shore establishments.

1.4 METHODS OF TEST CONSTRUCTION AND VALIDATION

1.4.1 Construction and Standardization of Group Tests (Chapter 14)

Test construction and validation are analyzed in detailed outline. The first step consists of discovering what critical abilities or skills are required by the job for which the test is being constructed. It is further necessary to acquire at least a general notion of the characteristics of the group to whom the test will be given. Also, the conditions and limitations of the classifications program as a whole necessarily affect the new test and must be considered in planning its details.

After this background work is completed, the test itself can be written in experimental form. Detailed statistical analysis of the responses made to each item by a sample group of men provides the information necessary to select items for the final test in terms of their difficulty and validity, to decide on time limits, to make certain that instructions are clear, and to arrange the items in proper order.

1.4.2 Contributions to Test Administration and Test Theory (Chapter 15)

An investigation of the accuracy with which aptitude tests were being scored in 1942 in the Navy revealed frequent and large errors. Methods of controlling and checking scoring procedures were devised. Subsequent investigations showed that errors were greatly reduced in frequency and in size.

Several innovations in test construction practice were tried out. They included:

1. A simple method of measuring the premium which a test places on speed by comparing the number of testees who attempt the last item of the test with the number attempting the first one.

2. A method of estimating the validity of a test for a school population without waiting for training and performance measures to accumulate after giving the test. Instead of waiting, the test is given to a graduating class and their test scores are correlated with final grades. Simultaneous administration of the test to a sample of entering students provides a check on the accuracy of the predicted validity coefficient by permitting comparison of regression weights for entering and graduating classes.

3. A method for increasing the validity of a test by selecting items in terms of their corre-

lation with an external criterion such as success at the school, instead of their correlation with the total-test score.

1.5 THE FUTURE OF RESEARCH IN MILITARY CLASSIFICATION

What remains to be done in the field of aptitude testing for military use? The experience of the Applied Psychology Panel suggests three principal needs.

1. New tests are necessary. Tests now available are not uniformly satisfactory. Better measures of certain types of ability need to be developed. In addition, changing equipment and changing methods increase or decrease the importance of careful selection of different kinds of specialists. Selection programs must keep up with these changes.
2. Psychological tests by their very nature require follow-up and review. They need maintenance and recalibration just as a physical measuring device requires maintenance and recalibration.
3. The actual usefulness of tests during World War II was considerably less than their potential usefulness. A number of administrative problems must be solved before the Army or the Navy can derive full benefit from their research programs on aptitude measurement and personnel classification.

These points are discussed below.

1.5.1 New Tests Are Necessary

Among the more general capacities for which better tests are needed are emotional fitness for combat, vocational interests, and visual and auditory capacity. Tests exist in each of these fields, but they are less effective than the general classification or mechanical aptitude tests.

The special abilities for which better tests are needed will vary from time to time depending upon the types of equipment to be used and the methods current for their use. For example, there is now reason to believe that the skill of the operator of a fire control director is relatively unimportant compared with the skill of the maintenance man responsible for calibration and alignment of the parts of the antiaircraft battery. In general, as equipment is more highly developed, it becomes possible to simplify some aspects of the operator's task. This simplification is frequently made at the expense of complicating the maintenance and calibration difficulties of the equipment. With such a change, a larger portion of the military population can serve satisfactorily as operators; their selection is therefore less critical. But the selection of maintenance personnel becomes more critical.

Because equipment changes, personnel requirements change. Because personnel requirements change, some tests become obsolete and others become necessary. New tests will be necessary as long as new equipment and new methods of warfare continue to be developed.

In developing new tests, more attention than was paid in World War II should be given to developing tests which are statistically independent of each other. The usefulness of a test in military classification depends, among other things, on its correlation with other tests also being used for classification purposes. Tests which have low correlations with other tests in a battery are more valuable than tests which have high correlations. Enough apt men for all specialties can be found only by having a test battery which picks out the special aptitudes of each man and furnishes a number of scores each of which is relatively unaffected by general ability.

In developing new tests, more emphasis than was given in World War II should be placed on field or combat skill as a criterion for validating the tests. Even at best, school grades are an unsatisfactory substitute for actual performance records in the validation of selection tests. School grades were frequently used in World War II because they were easily available and because they were more reliable than combat records. However, the ultimate goal of both selection and training is the achievement of highly developed combat skill. Improving the ability to predict that skill requires continuous effort to secure more reliable combat performance criteria.

1.5.2 Aptitude Tests Require Follow-up and Review

A test, like any precision instrument, must be adjusted and maintained as time passes. It cannot ordinarily be constructed, put to use, and allowed to continue in use without regular review and adjustment. An item which is satisfactory in difficulty when most recruits are coming from one section of the population may be too easy or too hard when the source of recruits changes. A change in the educational system of the country, for example an increase in the number of high school graduates, may make a formerly satisfactory test too easy. Susceptibility to these influences can be minimized, but it cannot be avoided completely. Only continuous watch over the tests and periodic statistical analysis of the distributions of item difficulties, test scores, and validity coefficients can keep them at maximal usefulness.

1.5.3 Administrative Use of Tests

During World War II several factors decreased the usefulness of the information obtained by aptitude tests. Two deserve special mention: the fallibility of interviewers' judgment, and ignorance regarding the usefulness and availability of test scores.

Interviewers' Judgment

Evidence was collected at the end of World War II which indicated that the work of the classification interviewer could be modified with resulting increase in the value of his recommendations. The evidence came from a direct comparison of the accuracy of interviewers in predicting training school performance with the accuracy of prediction based on test scores alone.[1] The test scores alone were clearly superior. The correlation between scores on a single test, the Mechanical Knowledge Test, and final grades of 3,496 men in electrician's mate schools was .50; the correlation between interviewers' recommendations and final grades of the same men in the same schools was .41.

In making his recommendations the interviewer had the test scores. He also had a good deal of additional information which he himself had collected during the interview. He should have been able to make better predictions using all this information than was possible with the scores on a single test. Evidently, most interviewers were too much impressed with the information they had secured and gave it greater weight than they assigned to the relatively more objective and reliable evidence of the test scores.

Better training of interviewers, the use of more objective and reliable methods of securing information obtained in the interview, and standardized weighting of the variables to be considered in selecting an appropriate assignment will make it possible for the interviewer's judgment to add to, rather than subtract from, the value of assignment by test scores alone. Methods for bringing about this improvement are described in Chapter 11.

Personnel Policies

Psychological tests are relatively new. Their military value, however, was demonstrated in World War I. They were widely used in World War II. Nevertheless, the administration of personnel policies, many of which grew up long ago, frequently rendered the test information useless. Two illustrations may be given.

The first illustration arises from the operation of the quota system in making assignments. Pools of men of all kinds are maintained at various Army and Navy centers. As new ships are commissioned, as new units are activated, or as new classes enter a training school, men are drawn from these pools. Frequently the pool is not tapped until a few days before the quota must be filled. It is then necessary to take the best men available at that time and in that pool. This sometimes means that men not qualified for school training must be included in order to fill the quota. Sometimes it means that a man highly trained in one specialty must be assigned to a totally new kind of duty because the pool is not large enough at the particular moment to fill the quota appropriately. As a specific example, the Army assigned a number of college graduates who had had special training for radar maintenance to duty

as truck drivers and as mess sergeants because there were more radar men and fewer truck drivers and mess sergeants than needed at that particular place and time.

The dates or approximate dates of commissioning a new ship, activating a new unit, or starting a new class are usually known for at least some weeks in advance. It is not easy to work out administrative procedures which will take advantage of this advance knowledge to minimize the number of misassignments resulting from the quota system. But advantage has occasionally been taken of it, and service efficiency was improved as a result.

A second illustration of failure to utilize carefully collected information about each man's abilities comes from an old shipboard practice. Men are frequently assigned to their bunks alphabetically, and that assignment is naturally one of the first things done for a man when he comes aboard ship. On shipboard men must have their bunks near their battle stations in order to meet surprise attack promptly. The result is that the position of a man's name in the alphabet may have more to do with determining his battle station than does all the information available about his abilities.

The total effect of administrative malpractice is unknown. The following quotation from an Army report, however, indicates that it is sometimes very large. "About March 1, 1945, the Armored Replacement Training Center completed training of its 150,000th replacement. Where these men have gone is one of the mysteries of the war. Certainly they have not gone to the units for which they were trained. Some units state that they have received no men with armored training; others comment upon the few that arrive."[2] To have most of 150,000 men, all specially trained for one type of duty, disappear into a variety of other assignments is evidence that there is still much to be done in the way of administration of sound personnel policies before maximal effect can be obtained from present personnel selection and training programs.

Better personnel procedures and better standardization of them will help overcome these difficulties. So will more stringent quality control policies. The greatest gain will be achieved when regular officers become as aware of the importance of getting the right man for a job as they are of getting the right ammunition for a gun.

Chapter 2

THE UNITED STATES NAVY BASIC CLASSIFICATION TEST BATTERY

By *Dael Wolfe*[a]

Summary

When expansion of the U. S. Navy began, it became apparent that the tests and procedures being used to classify and assign enlisted personnel were out of date and inefficient. A Test and Research Section was established in the Bureau of Naval Personnel to improve the classification tests. At about the same time, the Applied Psychology Panel of NDRC was asked to establish a project for "Research and Development of the Navy's Aptitude Testing Program."

The Test and Research Section and the NDRC project, working in close cooperation, developed the U. S. Navy Basic Classification Test Battery. After it proved successful, two additional forms of the battery were prepared for use ashore. A fleet edition of the General Classification Test was also developed.

Statistical analysis of the results of the new tests, in comparison with similar analyses of the tests in use before 1942, showed the new tests to be reliably superior to the old ones. Comparisons between test scores and grades made in a variety of Service schools showed that the tests predicted well which men would and which would not do well in different Service schools.

2.1 INTRODUCTION

For a number of years preceding the war all Navy recruits were given a set of paper-and-pencil tests which were intended to help in the selection or rejection of applicants and in the assignment of successful volunteers. This set of tests remained unchanged for many years. In 1942 the Navy decided that its test program should be brought up-to-date and that tests should be developed to aid in the tremendously increased assignment problems of a wartime Navy. A new and very small section was established in the Training Division, Bureau of Naval Personnel, for this purpose. At about the same time, the National Research Council's Committee on Service Personnel—Selection and Training (which later became the Applied Psychology Panel) was requested to make a survey of the procedures used in the selection, classification, and training of Navy personnel and to submit its findings and recommendations to the Bureau of Naval Personnel. The Bureau responded to this report[1] by an immediate request to OSRD to establish a project on "Research and Development of the Navy's Aptitude Testing Program." In response to this request, Project N-106 was formed under a contract with the College Entrance Examination Board, Princeton, New Jersey.

2.1.1 The Basic Classification Test Battery

Through the joint and always closely cooperative work of the rapidly expanded Test and Research Section of the Bureau of Naval Personnel and Applied Psychology Panel Project N-106, there was constructed a battery of aptitude tests designed to improve efficiency in the placement of Navy recruits and to assist in the maximum utilization of the Navy's manpower. These tests were collectively called the U. S. Navy Basic Classification Test Battery. The battery contains five tests printed in four test booklets. Those developed by the Applied Psychology Panel project are marked with an asterisk.

*1. General Classification Test (GCT). This is a composite test of 100 items, divided into three subtests as follows:

a. Sentence Completion subtest (30 items, 10 minutes).

[a] This chapter is based primarily on the work of the staff of NDRC Project N-106.

b. Opposites subtest (30 items, 8 minutes).

c. Analogies subtest (40 items, 15 minutes).

*2. Reading Test (30 items, 25 minutes).

*3. Arithmetical Reasoning Test (AR) (30 items, 30 minutes). (Both the Reading and Arithmetical Reasoning Tests are bound in the same booklet.)

4. Mechanical Aptitude Test (MAT). This is a composite test of 129 items, divided into three subtests as follows:

a. Block Counting subtest (45 items, 6 minutes).

b. Mechanical Comprehension subtest (44 items, 20 minutes).

c. Surface Development subtest (40 items, 8 minutes).

5. Mechanical Knowledge Test (MKT). This is a composite test of 135 items, of which 60 relate to electrical knowledge and 75 to mechanical. Both an electrical and a mechanical score are obtained from the test. The total time allowed is 37 minutes.

Responsibility for the construction of some of these tests was assigned to the project, while others were developed by the Bureau. Actually there was cooperative work on all tests. Detailed discussions and planning by both groups preceded the development of all the tests. Tests constructed by either group were criticized by the other. In many respects the Applied Psychology Panel's project on "Research and Development of the Navy Aptitude Testing Program" served as a closely integrated part of the Test and Research Section of the Bureau of Naval Personnel.

2.2 THE OLD NAVY TEST BATTERY

2.2.1 Statistical Analysis of the Old Battery[2, 4, 7, 8]

The first step in the construction of the Basic Classification Test Battery was a thorough study and statistical analysis of the tests formerly used for recruit selection and assignment.

These tests were:

1. The O'Rourke General Classification Test, Junior Grade, Form C and D, Navy Edition.
2. The O'Rourke Mechanical Aptitude Test.
3. U. S. Navy Standard Recruit Test in Arithmetic.
4. U. S. Navy Standard Recruit Test in English.
5. U. S. Navy Standard Recruit Test in Spelling.
6. U. S. Navy Radio Code Aptitude Test.

The Bureau of Naval Personnel supplied the Applied Psychology Panel with answer sheets for these six tests from 43,539 recruits. The tests were given at all naval training stations which existed at that time. These stations were located at Great Lakes, Illinois; Newport, Rhode Island; San Diego, California; and Norfolk, Virginia. All the men had been tested between January and April 1942.

A sample of these 43,539 papers was carefully rescored by the project staff and then submitted to detailed statistical analysis to determine the difficulty, reliability, and internal characteristics of the tests. Correlations with Service school grades were also computed to

TABLE 1. Reliability coefficients of old Navy aptitude tests.

Training Station	GCT (Form B)	GCT (Form C)	Code	Mechanical Aptitude Test	Arithmetic	English	Spelling
Newport	.95	.95	*	.94	.89	.91	.94
Norfolk	.95	*	.69	.94	.88	.89	.93
Great Lakes	*	.94	.75	.94	.87	.86	.92
San Diego	*	.96	.78	.95	.88	.92	.92
Number of items	100	100	78	162	20	100	50
Time allowed (in min)	60	60	...	55	25	30	20

*Information not available.

determine the validity of the tests. Samples of the papers from different naval training stations were compared for the information such comparisons could give on the Navy population. This work was all done as a background for the development of the U. S. Navy Basic Classification Test Battery and to help in the necessary decisions as to the length, difficulty, type of item, etc., which should characterize the tests in the new battery.

Difficulty[4]

The frequency distributions of scores indicated that the General Classification, English, and Spelling Tests were too easy to provide the most efficient differentiation. The Arithmetic and Mechanical Aptitude Tests were of suitable average difficulty but failed to exhibit as great a concentration of scores in the middle range as would be desirable.

With respect to item difficulty, the ideal form for the distribution of the difficulty of the items of a general purpose test is not known exactly. It seemed, however, that for tests whose items have a comparatively high correlation with the total test score, a distribution of item difficulties should be flatter than for tests with low item-test correlation. This relation was not found to hold in the tests under investigation.

Reliability[2]

The odd-even reliability, corrected by the Spearman-Brown formula, of the individual tests making up the old Navy Battery was satisfactory for all tests except the test of code aptitude. The reliability coefficients for each test as determined at the four training stations are presented in Table 1.

The two parts of the Mechanical Aptitude Test were each reliable, but the correlation between them was only about .65 to .70, which made questionable the desirability of combining the two parts into a single test score.

Intercorrelations[4]

The intercorrelations of five of the tests, omitting the Radio Code Aptitude Test, are given in Table 2. Some of the intercorrelations were high enough to suggest the desirability of omitting one or two of the tests.

Item Analysis[7]

The detailed item analysis of the tests gave information on the usefulness of the individual items.

Of the 11 different types of items contained in the General Classification Test, the five types with the highest average biserial r's were, in descending order, proverbs, completion, opposites, synonyms, and analogies. This finding

TABLE 2. Intercorrelations of old Navy aptitude tests ($N=500$ for each station).

	GCT	MAT	Arithmetic	English	Spelling	Average
GCT						
Newport	...	.46	.51	.58	.58	.53
Norfolk	...	.50	.61	.64	.60	.59
Great Lakes	...	.42	.49	.54	.53	.50
San Diego	...	.63	.56	.69	.66	.64
MAT						
Newport		...	.29	.29	.22	.32
Norfolk		...	.39	.32	.30	.38
Great Lakes		...	.27	.29	.22	.30
San Diego		...	.47	.53	.39	.51
Arithmetic						
Newport			...	.46	.44	.43
Norfolk			...	.49	.49	.50
Great Lakes			...	.48	.52	.44
San Diego			...	.57	.55	.54
English						
Newport				...	.54	.47
Norfolk				...	.55	.50
Great Lakes				...	.56	.47
San Diego				...	.59	.60
Spelling						
Newport					...	.45
Norfolk					...	.49
Great Lakes					...	.46
San Diego					...	.55

supported the view that items which definitely require some reflection or cogitation are, in general, superior to those which can be answered with little or no thinking.

A careful study was made of items which showed exceptionally high item-test correlations at each station and those which showed exceptionally low item-test correlations. The detailed earmarks of good versus poor items were found not to be uniform from one test to another. It appeared that the reasons for a low biserial correlation with total score were gen-

[redacted]

erally more discernible than the reasons for a high correlation. For practical work in test construction this finding suggests that the very poor items of a test may often be caught and rejected simply by careful inspection without the need for an actual tryout. The opposite, however, was not true. No consistent characteristics of exceptionally good items could be found. Probably the most important general conclusion from this part of the study is that no prediction, however sagacious, can substitute adequately for the detailed quantitative information about an item that is yielded by an actual trial. The actual trial of items is especially useful for detecting excessively close or excessively remote (nonfunctional) distractors and for the identification of the most effective items in a test.

VALIDITY[8]

In order to determine the validity of the tests comprising the old battery, the scores on each of the six tests were correlated with grades in six Service schools at the Great Lakes Naval Training Station. The Service schools were electrician's mate, fire controlman, gunners mate, torpedoman, quartermaster, and signalman. The principal results of the validity study were:

1. The Service school grades used as criteria were highly reliable. The reliability coefficients varied from .93 to .97. These coefficients were determined by correlating the sum of the grades for odd-numbered weeks with the sum of the grades for even-numbered weeks and then applying the Spearman-Brown correction formula.

2. The validity of the individual tests was generally low. Only five of the 36 validity coefficients were as high as .29. These five were .54, .38, .37, .37, and .29.

3. The General Classification Test was a better predictor of school grades than any of the more specialized tests.

TABLE 3. Summary of data on validity of old Navy aptitude tests.

	Group I								Group II			
	EM		FC		GM		TM		QM		SM	
	Tests	Corr.	Tests	Corr.	Tests	Corr.	Tests	Corr.	Tests	Corr.	Tests	Corr.
Best single test (r)	GCT	.54	Arith.	.38	Arith.	.27	MAT	.24	Arith. Spell.	.27 .27	GCT	.31
Best single test (r_c)	GCT	.72	GCT	.65	Arith.	.42	MAT	.28	Arith.	.40	GCT	.51
Best single variable (r)	GCT	.54	Arith.	.38	Age	.28	MAT	.24	Arith. Spell.	.27 .27	GCT	.31
Best combination of two tests (R)	GCT-MAT GCT-Arith. GCT-Eng.	.55 .55 .55	GCT-Arith.	.44	Arith.-GCT Arith.-MAT	.32 .32	MAT-RCA	.26	RCA-Arith. RCA-Spell.	.34 .34	GCT-RCA	.36
Best combination of two variables (R)	GCT-MAT GCT-Arith. GCT-Eng. GCT-Age	.55 .55 .55 .55	Arith.-Age	.46	GCT-Age	.37	MAT-Age	.31	Arith.-Age	.36	GCT-RCA	.36
Best combination of three tests (R)	GCT Eng. MAT GCT Eng. Arith	.56 .56	GCT Arith. MAT	.46	GCT Arith. MAT	.34	GCT MAT RCA	.27	Arith. Spell. RCA	.38	GCT Spell. RCA	.41
Best combination of three variables (R)	GCT Eng. Age	.57	GCT Arith. Age	.51	GCT Arith. Age	.40	GCT MAT Age	.31	Arith. Spell. Age	.39	GCT Spell. RCA	.41

4. Best prediction was found in those schools which contained a large portion of academic work.

5. The best combination of two tests gave multiple correlations which on the average were only .04 higher than the best single test.

The chief statistical results of the validity study are summarized in Table 3. The values of r_c in the second line of the body of this table are the correlations corrected for curtailment of the distribution of the sample studied.

The tests included in the new Basic Aptitude Test Battery and in supplementary Navy tests contain items of every type which this validity study suggested might be of value.

TABLE 4. Means and standard deviations of Navy aptitude test scores for four stations (based on 43,000 cards).

Test by stations	Number of cases	Mean	Standard deviation
GCT			
Newport	4,858	69.35	14.53
Norfolk	11,740	65.72	20.59
Great Lakes	15,878	77.02	13.16
San Diego	8,489	66.85	17.39
MAT			
Newport	4,837	46.25	15.31
Norfolk	10,798	33.83	25.86
Great Lakes	15,650	55.04	27.27
San Diego	7,866	51.54	16.82
Arithmetic			
Newport	4,864	48.17	23.89
Norfolk	11,753	44.75	26.80
Great Lakes	15,820	65.03	23.45
San Diego	7,808	47.83	25.00
English			
Newport	4,852	65.81	12.63
Norfolk	11,752	33.35	21.56
Great Lakes	15,831	71.32	11.65
San Diego	7,092	63.86	15.23
Spelling			
Newport	4,832	57.13	20.57
Norfolk	11,739	51.91	24.58
Great Lakes	15,199	67.71	19.26
San Diego	7,068	57.82	21.23

VARIABILITY IN NAVY SAMPLES[2, 4]

Consistent differences were found in the average scores and in the distributions of scores reported from the four naval training stations from which answer sheets were obtained. On all tests the highest average scores were made by the recruits at the Great Lakes Naval Training Station. The lowest average scores were all made by recruits at the Norfolk Naval Training Station. The data are presented in Table 4.

Comparisons based on a much smaller group of papers, which had been very carefully rescored by the project, showed smaller differences between the four naval training stations but did not alter the rank order. The number of cases involved was from 200 to 300 for each test from each station. The smaller population was selected at random from all the papers available from each station. In this smaller and more carefully scored sample the average score at Norfolk was generally from one third to one half a standard deviation below the Great Lakes average.

Discrepancies in the cards on which the scores were recorded and gross differences between the reports coming from the different stations suggested that differences in examination and scoring procedures were large and made it appear probable that testing conditions had sometimes departed seriously from recommended procedures.

2.2.2 Revision of Mechanical Aptitude Test

Upon request of the Standards and Curriculum Division, Bureau of Naval Personnel, detailed suggestions were made for the revision of the mechanical aptitude test then in use in the Navy.[5] These suggestions were accepted and incorporated into new forms of a mechanical aptitude test which the Standards and Curriculum Division constructed as part of the Basic Battery.

2.3 EXPERIMENTAL TESTS

In order to secure additional information on the types of tests which should be included in the basic classification test battery, an experimental battery of 11 tests was administered to entering classes in Gunners Mate, Fire Control-

[redacted]

man, and Rangefinder Operator's Schools at the Navy Yard, Washington, D. C.[10] The 11 tests included a verbal test, tests of quantitative ability, spatial relations tests, and a test of mechanical aptitude. After the men had graduated, their Service school grades were used as criteria in evaluating the 11 tests.

As judged by predictive efficiency, the seven best tests turned out to be tests of arithmetical reasoning, arithmetical computation, vocabulary, mechanical comprehension, surface development, the old general classification test, and the old mechanical aptitude test. The new Navy Basic Classification Test Battery was designed to contain items similar to those in all these tests except arithmetical computation (see Section 2.4). A new test of arithmetical computation was prepared later by the project (see Section 2.5).

2.4 CONSTRUCTION OF TESTS FOR THE NEW BATTERY

A new General Classification Test, a new Reading Test, and a new test of Arithmetical Reasoning were constructed by Applied Psychology Panel Project N-106[3] for inclusion in the new Basic Classification Test Battery.

Certain specifications for these new tests were stated by the Bureau of Naval Personnel. These were:

The general classification test was to include only items involving verbal ability. Types of items suggested were sentence completion, opposites, and analogies.

A sentence completion item is a sentence with one word omitted; the task is to choose the one of several words presented which best fits in with all parts of the sentence. In such items, it should be necessary for the candidate to sense the implications of all parts of the sentence in order to choose the correct word. The choice should not depend simply on vocabulary knowledge, nor should the test measure good word usage in a purely literary sense.

An opposites item consists of a key word followed by several other words, one of which is the opposite of the key word. The task is simply to choose the word which means the opposite of the key word. Highly technical words were to be avoided. Words likely to occur in Navy manuals were recommended for use.

An analogies item is a statement of an analogy with the last word missing. The task is to choose, from a list of several words, the one which best completes the analogy. Items of this type should depend upon general verbal ability rather than upon vocabulary or special information.

The reading test was to consist of paragraphs of increasing difficulty, each followed by multiple-choice questions on the content of the paragraph. It was suggested that the questions should measure ability to draw inferences, to note detail, and to follow directions.

The arithmetical reasoning test was to consist of verbally stated problems which involve ability in quantitative thinking. The statement of the problems should be sufficiently simple that reading comprehension is not an important factor and that the actual reading of the problems consumes a small proportion of the testing time. The problems should be capable of solution by simple arithmetic, and the computations should be simple enough that a relatively small proportion of the testing time is spent in actually multiplying, dividing, etc. In other words, half or more of the testing time should be spent in thinking out the method of solution.

The general classification test which was prepared included 30 sentence completion, 30 opposites, and 40 analogies items, making a total of 100 items. All were multiple-choice items with 5 choices. The reading test was made up of 6 paragraphs with an average of 5 multiple-choice questions on each, making a total of 30 items. The arithmetical reasoning test also contained 30 five-choice items. These numbers were consistent with time limits suggested by the Bureau. The items making up each part were arranged in estimated order of difficulty. The attempt was made to make the range of difficulty of the items so great that the easiest item could be answered correctly by approximately 95 per cent of the recruits and the most difficult items could be answered correctly by about 5 per cent.

The General Classification Test, Reading

[redacted]

Test, and Test of Arithmetical Reasoning are reproduced in the appendix as they were printed by the Navy from copy prepared by Applied Psychology Panel Project N-106. An unclassified account of the preparation of the Basic Battery tests (Form 1) has been published by the Navy.[20]

2.4.1 Statistical Analysis of the Basic Battery (Form 1)

A national sample of 400 answer sheets for each test in the Basic Classification Test Battery (Form 1) was analyzed in detail in order to answer a number of questions concerning the characteristics of the tests making up the battery. The questions are of general interest and can appropriately be asked of any set of tests. Since the answers were specific to Form 1 of the Basic Battery they are not given here in detail. Full information is contained in the original report.[14] The specific questions are given below.

INTERNAL CHARACTERISTICS OF A GOOD TEST BATTERY

1. Does each test and subtest yield a suitable distribution of scores?
    a. Is the range or spread of scores adequate?
    b. Are the distributions of scores free from significant skewness?
    c. Are the distributions reasonably similar in shape, so that standard scores for different tests may be considered comparable in meaning?
    d. Is each test and subtest of suitable difficulty (neither too easy nor too hard)?

2. Are the tests adequately reliable for use in individual placement?

3. Are the correlations among tests sufficiently low to provide measurement free from excessive overlapping?

4. Are the correlations among the subtests of the Mechanical Aptitude Test sufficiently high to justify combining the subtest scores into a single Mechanical Aptitude Test score? Similarly, are the correlations among the subtests of the General Classification Test sufficiently high?

5. Do any of the component subtests of the General Classification Test or the Mechanical Aptitude Test have an unduly large or small influence in determining total-test scores?

6. Do differences exist among the various tests or subtests in the premium granted for speed of performance? Should the time limits for any of the tests or subtests be changed?

7. Concerning the individual items:
    a. Is there a proper distribution of item difficulty in each test or subtest? More specifically, is the proportion of very easy and very difficult items sufficiently low, and is the dispersion or spread of difficulty satisfactorily adjusted to the average intercorrelation among items in the test or subtest?
    b. Are the items in each test or subtest arranged in order of difficulty?
    c. Are the items within each particular test or subtest adequately homogeneous—i.e., do they succeed reasonably well in measuring the same function?
    d. Which particular items in each test or subtest should be rejected or revised, because of lack of homogeneity with their fellow items?
    e. Do the difficult items tend to reduce the homogeneity or internal consistency of any of the tests or subtests of the Basic Battery?

CHARACTERISTICS OF FORM 1 TESTS[14]

Distribution of Scores. The distributions of test scores for the Form 1 tests were, in general, satisfactory. Most of the tests were a little too difficult. One, the Surface Development subtest of the Mechanical Aptitude Test, did not differentiate well among either the very high or the very low men. The easiest questions were not easy enough and the most difficult ones were not hard enough. In consequence, there was a tendency for scores to pile up at both ends of the distribution. Only three of the subtests showed any statistically significant degree of skewness. Likewise three departed from normal in terms of kurtosis; scores on the electrical-pictorial items showed an excessive peaking while the distributions for Surface Development and Block Counting were excessively flat.

Reliabilities. Odd-even reliabilities, corrected by the Spearman-Brown formula, were as shown in Table 5. All reliability coefficients were based on a random selection of 200 cases from the basic sample of 500. The reliability coefficient for the mechanical aptitude scores was, to some extent, spuriously high, since speed of performance played a considerable role in determining these scores, and it is well known that placing a premium on speed tends to raise the reliability coefficient obtained by the odd-even technique. This was particularly true of the Block Counting subtest of the MAT.

TABLE 5. Reliabilities of Basic Battery tests.

General Classification Test		.94
Sentence Completion	.85	
Opposites	.84	
Analogies	.84	
Reading		.85
Arithmetical Reasoning		.83
Mechanical Aptitude		.97
Block Counting	.95	
Mechanical Comprehension	.84	
Surface Development	.98	
Mechanical Knowledge, Electrical		.89
Mechanical Knowledge, Mechanical		.90

Intercorrelations. The correlations among the 12 tests or subtests are shown in Table 6. Most of these correlations were satisfactory, but they suggested several possible revisions of tests or procedures. The three intercorrelations among the subtests of the MAT were .48, .51, and .60. There is an inconsistency in combining these into a single MAT score while reporting separately the GCT and Reading scores which correlated .81 with each other.

Speed. Several of the tests or subtests, particularly Surface Development, but also Sentence Completion, Arithmetical Reasoning, and Mechanical Comprehension, put more premium on speed than had originally been intended. Revised time limits were recommended. A method of measuring the premium placed on speed in each test is given in Section 15.3.

Item Arrangement in Terms of Difficulty. In general the easy items on any test were found in the early part of the test and the hard items toward the end, but in none of the tests was there a smooth and gradual progression from easy to hard items. Rearrangement of the items would have improved every one of the tests in this respect.

Item Difficulty and Validity. Information on the difficulty of each item and the biserial correlation of each item with the subtest of which it was a part are given in reference 14.

The detailed information obtained in this analysis of the Form 1 tests confirmed the judgment of the Bureau of Naval Personnel, which had put the tests into use as soon as they were prepared and without waiting for extensive field trials. Original preparation was based on an extensive background of experience in test construction, and it was, therefore, expected that the tests would work well. When the statistical results became available, they fully justified that expectation.

TABLE 6. Intercorrelations among tests* and subtests† of Basic Battery (national sample, $N = 500$).

	GCT	Sent. Comp.	Opp.	Anal.	READ.	ARITH. REAS.	MECH. APT.	Block Count.	Mech. Comp.	Surf. Devel.	M.K. ELECT.	M.K. MECH.
GCT	...	.89	.90	.90	.81	.69	.60	.46	.57	.48	.53	.49
Sent. Comp.		...	.74	.69	.73	.60	.55	.42	.48	.47	.46	.42
Opposites			...	.71	.73	.61	.47	.38	.46	.36	.48	.46
Analogies				...	.74	.64	.59	.44	.59	.47	.50	.44
READING					..	.69	.56	.44	.52	.47	.51	.46
ARITH. REAS.						...	.61	.52	.51	.51	.47	.41
MECH. APT.							...	.87	.74	.86	.53	.55
Block Count.								...	.51	.60	.42	.45
Mech. Comp.									...	.48	.60	.62
Surf. Devel.										...	.36	.35
M.K., ELECT.											...	.78
M.K., MECHAN.												...

*Written in CAPITAL letters.
†Written in small letters.

[redacted]

2.4.2 The Basic Classification Test Battery, Form 2

When the Bureau of Naval Personnel decided to construct a new form of the Basic Classification Test Battery, Project N-106 was asked to develop new forms of the three tests which it had originally prepared—the General Classification Test and the tests of Reading and Arithmetical Reasoning.[11]

The procedure followed in constructing the Form 2 tests involved four major steps.

1. An experimental edition of each of the three tests was first constructed. An effort was made to match the items of Form 1 in content and difficulty, but in order to allow some selection of items each part of the experimental edition contained about 30 per cent more items than were planned for retention in the final edition.

2. These experimental tests, along with Form 1 of the same tests, were administered to recruits at the U. S. Naval Training Station at Sampson, New York.

3. The items of both forms of each test were statistically evaluated. Information gained in the detailed analysis of Form 1 (see Section 2.4.1) helped here to improve Form 2.

4. Finally, the items to be included in the new Form 2 of each test were selected on the basis of information gained in these item evaluations. The individual items in Form 2 of each test were selected on the basis of three criteria: (1) the difficulty of the items in each test; (2) the correlation with total test score; and (3) their comparability with the individual items of the Form 1 tests.

By this procedure a second form of each test was constructed which contained items satisfactorily distributed in terms of difficulty, highly correlated with total-test or subtest scores, and comparable to the items in the parallel Form 1 test.

A more detailed description of the procedures used in constructing Form 2 of the U. S. Navy General Classification Test and the U. S. Navy Tests of Reading and Arithmetical Reasoning is contained in Chapter 14, which describes the test construction procedures recommended and used by this project.

Form 2 of the General Classification Test, Reading Test, and Arithmetical Reasoning Test were put into regular use by the Navy as soon as the Form 2 tests were all completed. The three tests constructed by Project N-106 are reproduced in the appendix.

2.4.3 The Fleet Edition of the General Classification Test

Forms 1 and 2 of each of the tests in the Basic Classification Test Battery were designed to be administered by competent classification specialists or experienced officers. Later it appeared desirable to have available a self-administering edition of the General Classification Test which could be used in the fleet where classification specialists and test scoring machines were not available. Consequently, the General Classification Test, Fleet Edition, was constructed.[12]

The item content of the Fleet Edition is identical with Form 1 of the GCT. (Form 1 is reproduced in the appendix.) The two forms differ in that the Fleet Edition has modified instructions, an overall time limit instead of separate time limits for the three parts, and does not require a separate answer sheet.

Norms for the Fleet Edition and tables for converting raw scores into Navy standard scores (which always have a mean of 50 and a standard deviation of 10) were obtained by giving Forms 1 and 2 to one group of subjects and giving the Fleet Edition and Form 2 to another comparable group of subjects. By this procedure Form 2 served as a bridge connecting Form 1 and the Fleet Edition and made possible the establishment of norms and conversion tables. Direct comparison between Form 1 and the Fleet Edition was not possible; since the item content of the two was identical, conversion tables could not be constructed by giving both forms to the same group.

2.4.4 Other Studies of Tests in the Basic Classification Test Battery

The Bureau of Naval Personnel requested a statistical analysis of Form 1 of the Mechanical Knowledge Test, which had been con-

[redacted]

structed in the Bureau.[9] The data available for analysis consisted of 230 answer sheets for the test, administered in April 1943, at the Bainbridge Naval Training Station for purposes of preliminary standardization.

The Mechanical Knowledge Test consisted of 135 items. Of these, 60 were designed to yield an electrical score and 75 to yield a mechanical score. Both the electrical and mechanical scores were based on two types of items, pictorial and verbal.

The reliability of the total test and the reliability of the mechanical knowledge portion of the test were both found to be .91, which is satisfactory for individual use. The reliability of the electrical score was found to be .79, hardly satisfactory for individual prediction. All reliability coefficients are odd-even correlations corrected by the Spearman-Brown formula.

The correlation between electrical and mechanical scores, .73, was high enough to justify combining the two, instead of reporting separate scores on the two parts of the test.

Differences between the electrical scores and mechanical scores, obtained by the men in the sample studied, were calculated to have a reliability (odd-even, Spearman-Brown corrected) of .44. With such low reliability, differences between the two scores would have to be relatively large before they could be used with confidence in recommending a man for assignment to a school or billet which required one kind of knowledge (mechanical or electrical) to a considerably greater extent than it required the other. It was therefore recommended to the Bureau of Naval Personnel that only differences of 11 points or more in Navy standard score should be given serious consideration in making such differential assignments.

Factor Analysis of the Basic Classification Test Battery[6]

The Basic Classification Test Battery was designed to measure four types of ability: (a) verbal ability, by the General Classification Test and the Reading Test; (b) quantitative reasoning ability, by the Arithmetical Reasoning Test; (c) mechanical knowledge, by the Mechanical Knowledge Test; and (d) aptitude in dealing with spatial and mechanical relationships, by the Mechanical Aptitude Test.

Method of Analysis. The method of multiple factor analysis is appropriate to determine how distinct these four kinds of ability are as measured by the tests of the basic battery. Since several of the tests consisted of parts which were scored separately, it was possible to obtain 12 different scores, to find the correlations among these 12 scores, and to analyze the intercorrelations of the 12 scores by factorial methods.

The different scores which were used in the factor analysis were:

1. Reading = total score on the Reading Test.

Table 7. Reliabilities and intercorrelations of test scores used in factor analysis of the Basic Battery.

	1	2	3	4	5	6	7	8	9	10	11	12
1. Reading	.82*											
2. Opposites	.70	.86*										
3. Analogies	.73	.72	.88*									
4. Sent. comp.	.68	.70	.69	.85*								
5. Arith. reas.	.69	.61	.68	.61	.86*							
6. Tool rel.	.25	.24	.22	.18	.21	.80*						
7. Mech. inf.	.60	.57	.57	.52	.60	.61	.88*					
8. Elect. score	.50	.48	.49	.37	.49	.74†	.84†	.84*				
9. Mech. score	.51	.46	.44	.47	.49	.79†	.89†	.73	.90*			
10. Block count.	.31	.23	.32	.30	.43	.22	.34	.27	.31	.90*		
11. Mech. comp.	.45	.41	.42	.47	.55	.51	.64	.56	.62	.54	.81*	
12. Surf. devel.	.57	.43	.56	.49	.51	.35	.56	.46	.53	.44	.52	.93*

*These reliabilities were used for estimating specificity but were not used in factor analysis.

†These correlations were not used in the factor analysis.

[redacted]

2. Opposites = score on the Opposites subtest of the GCT.

3. Analogies = score on the Analogies subtest of the GCT.

TABLE 8. Test projections on the centroid axes.

Tests	Centroid axes		
	I	II	III
1. Reading	+.804	+.294	+.051
2. Opposites	+.750	+.364	+.182
3. Analogies	+.794	+.337	+.040
4. Sentence Completion	+.752	+.328	−.056
5. Arithmetical Reasoning	+.785	+.187	−.142
6. Tool relationships	+.479	−.464	+.353
7. Mechanical Information	+.795	−.231	+.296
8. Electrical score	+.700	−.281	+.408
9. Mechanical score	+.729	−.349	+.284
10. Block Counting	+.517	−.246	−.403
11. Mechanical Comprehension	+.725	−.368	−.169
12. Surface Development	+.704	−.126	−.050

4. Sentence completion = score on the Sentence Completion subtest of the GCT.

5. Arithmetical reasoning = score on the Arithmetical Reasoning Test.

6. Tool relationships = score on the pictorial section of the MKT.

7. Mechanical information = score on the verbal section of the MKT.

8. Electrical score = score on the electrical sections of the MKT.

9. Mechanical score = score on the mechanical sections of the MKT.

10. Block counting = score on the Block Counting subtest of the MAT.

11. Mechanical comprehension = score on the Mechanical Comprehension subtest of the MAT.

12. Surface development = score on the Surface Development subtest of the MAT.

Table 7 displays the correlations among these 12 scores. This table was furnished to the project by the Bureau of Naval Personnel.

Factorial Pattern. A centroid factor analysis, by the methods devised by Thurstone, gave the factor pattern shown in Table 8. The factor loadings after rotation are shown in Table 9.

The major conclusions of the factor analysis study were:

1. Three factors were found. One factor (Factor A) was clearly a verbal factor; the second (Factor B) a mechanical factor; and the third (Factor C) was not well defined in the analysis. These factors were relatively independent of each other, the highest intercorrelation (that between B and C) being .31 and the lowest (between A and B) being .19.

2. Verbal ability is well represented in the test battery. The Reading Test, the General Classification Test, and the Arithmetical Reasoning Test all represent this factor. Later validity studies (see Table 11 and, especially, Table 13) showed little consistent differential validity for the three tests in different Service schools. The possibility of pooling scores or of eliminating one of the tests, probably the Reading Test, should be considered.

3. An independent quantitative reasoning factor is not represented in the Arithmetical Reasoning Test. Other tests of a quantitative reasoning type might be tried to see if they would give scores which are more differentiated from verbal ability and which also measure an ability necessary for success in certain Service schools (see Section 2.5).

TABLE 9. Factor loadings after rotation into a simple structure and square root of specificity of tests.

Test	Factor A	Factor B	Factor C	Square root of specificity
1. Reading	+.704†	+.078	+.151	+.28
2. Opposites	+.758†	+.107	−.005	+.36
3. Analogies	+.730†	+.040	+.144	+.36
4. Sentence Completion	+.677†	−.032	+.217*	+.41
5. Arithmetical Reasoning	+.565†	+.011	+.347*	+.44
6. Tool Relationships	−.010	+.692†	−.003	+.48
7. Mechanical Information	+.342*	+.592†	+.089	+.33
8. Electrical Score	+.274*	+.674†	−.029	+.32
9. Mechanical Score	+.208*	+.642†	+.112	+.41
10. Block Counting	+.011	+.038	+.614†	+.64
11. Mechanical Comprehension	+.088	+.341*	+.517†	+.35
12. Surface Development	+.294*	+.257*	+.332*	+.65

*Factor loadings between .20 and .35.
†Factor loadings greater than .50.

4. An independent mechanical ability appears to be best represented by the tool relationships part of the Mechanical Knowledge Test. The two scores now reported, electrical

[redacted]

and mechanical, are very highly correlated and have essentially the same factor composition. If it is desirable to distinguish electrical from mechanical competence, then further work is needed on this differentiation. It may be possible, however, to emphasize the tool relationships section of the Mechanical Knowledge Test and so get a score that is relatively independent of verbal ability.

5. The block counting and mechanical comprehension sections of the Mechanical Aptitude Test have relatively high loadings on Factor C. The indications from the present battery are not very clear as to the nature of this factor. It may represent speed, spatial reasoning, quantitative reasoning, or some combination of these. A further investigation with additional tests would be necessary to determine what this factor represents.

2.5 THE ARITHMETICAL COMPUTATION TEST

2.5.1 Construction

Correlations between the several tests of the Basic Battery and the factor analysis study just reviewed showed the Arithmetical Reasoning Test to be more closely related to the General Classification Test than was considered desirable.[6, 14] The correlations between the individual part scores of the GCT and the Arithmetical Reasoning Test (AR) were all found to be above .60. In view of these findings, a purer test of *quantitative* ability was requested by the Bureau of Naval Personnel. It was desired to have a test which would depend less upon verbal ability than the Arithmetical Reasoning Test apparently did, but which would measure those types of computational skill demanded by Service school curricula. The Arithmetical Computation Test (AC) was developed to meet this request.[13, 15]

The first step in the development of the new Arithmetical Computation Test was the construction of a 60-item, free answer test.[13] Problems varied in difficulty from adding 56 and 25 to multiplying 2.96 by $.42\frac{2}{5}$. Problems of addition, subtraction, multiplication, and division of whole numbers, fractions, decimals, and mixed numbers were included.

The first experimental form of the new test, Form X-1, was administered to 1,430 recruits. The difficulty of each item was determined and used as a basis for selecting the items to be included in the second experimental form, Form X-2. The wrong answers most frequently given in answering the questions on Form X-1 were used as distractors in Form X-2.

Form X-2 consisted of 50 multiple-choice items. Six, seven, or eight alternative answers were supplied for each item. To prevent the men from choosing an approximate answer without working the problem through to a solution, the alternative "Answer not given" was included for each item; for seven items it was the correct alternative.

Form X-2 was administered to over 1,000 recruits at two Naval Training Stations. A sample of 500 papers, half from each station, was analyzed to determine the difficulty and the discriminative value of each item.

Form X-2 was found to have a very good distribution of item difficulties. No item was answered correctly by less than 20 per cent or more than 90 per cent of the group. The median was 61 per cent and there was a piling up of items in the 60 to 70 per cent range. This is a desirable feature since it represents a concentration of items where they can make the greatest discrimination.

The discriminative value of the items in Form X-2 was very high. The median biserial correlation with total-test score was .695.

The Arithmetical Computation Test, Form 1, the final form of the test, was constructed on the basis of the information gained in the analysis of Form X-2. It contains 30 items distributed nearly evenly (6 or 7 each) over the 4 fundamental processes of arithmetic. Nine involve whole numbers; 11, fractions; 6, decimals, and 1 both fractions and decimals. There are also 3 problems on percentages.

The difficulty range is from 28 to 87 per cent correct for 29 of the items. One very easy item was added at the start of the test. The biserial correlations with total score range from .40 to .84 with a median of .74.

Five alternatives were chosen for each item.

These were, generally, those most frequently selected by the group taking Form X-2. The distractors used on the final form were those chosen by men who made a lower total score in Form X-2 than the men choosing the correct alternative. "Answer not given" was included as an alternative for each item.

The odd-even, and Spearman-Brown corrected, reliability of the final 30-item test, estimated from the data collected on Form X-2, was .92.

2.5.2 Validation

A preliminary study[13] of the validity of the Arithmetical Computation Test for predicting grades in a number of different types of Service schools was made by correlating the scores of men who took the test at the time of their graduation from Service school with their Service school grades. The twelve validity coefficients for the Arithmetical Computation Test were found to range from .33 in aviation metalsmith school to .69 in basic engineering school. In six of the ten schools where comparisons could be made, the test had a validity which was as high as, or higher than, any other single Basic Battery test. Such validity coefficients indicate that the Arithmetical Computation Test is a good test if the assumption is correct that the training received in school prior to the test administration does not have a differential effect on performance. Evidence obtained later indicated that this assumption was correct. The evidence also indicated that, although the superiority of the AC over the other tests was not as great as was first reported, the AC showed a higher median validity coefficient (.43) than any test in the battery.[15]

The Arithmetical Computation Test was found to have high reliability and satisfactory validity. For a number of Service schools it proved to be more valid than any test in the Basic Battery. Yet it failed to achieve the original purpose of developing a valid test which would show appreciably lower correlations with the verbal tests of the Basic Battery than did the Arithmetical Reasoning Test. The correlations of Form X-2 with the GCT and the Arithmetical Reasoning Test for a recruit sample are shown in Table 10. The correlations would undoubtedly be considerably lower for a sample of Service school men than for this sample of recruits. Nevertheless the correlation with GCT was disappointingly high.

TABLE 10. Intercorrelations of GCT, AR, and AC tests.

Test	GCT (Form 2)	AR (Form 2)
AC (Form X-2)	.758	.782
GCT (Form 2)		.820

In constructing the AC test a very careful attempt was made to prepare a test which would not show a high correlation with verbal tests. The failure of this attempt suggests that it may be impossible to construct a computation test which will not be correlated with verbal measures when a wide range of talent is represented in the sample.

2.5.3 Use of AC Test

The new AC Test could be used in any one of four ways: (1) it might be substituted for the AR and thus become part of the Basic Battery; (2) the AR could be retained and scores on both arithmetic tests reported; (3) the AR and AC might both be used, but only the sum of the two scores reported; or (4) the AC could be used as a preliminary screening test and a more difficult arithmetical reasoning test be used as part of a battery of secondary tests.

The fourth possibility was recommended to the Bureau of Naval Personnel. If a secondary testing program should be instituted, the AC might well be used as a part of the primary testing battery; and the AR, in a more difficult form, in the secondary battery. Because of its high correlation with GCT, the AC Test could be incorporated into the General Classification Test and thus become a part of a general ability measure. (Chapter 12 contains a discussion of the possible advantages of using both a preliminary screening battery of tests and a second battery for use in making assignments to individual schools or billets.)

[redacted]

2.6 GENERAL APPRAISAL OF THE BASIC CLASSIFICATION TEST BATTERY

2.6.1 Abilities Measured by the Basic Battery

In developing the tests of the Basic Classification Test Battery, the Standards and Curriculum Division of the Bureau of Naval Personnel and the Applied Psychology Panel agreed on three major areas of human ability to be included—verbal ability, mechanical ability, and quantitative ability. The influence of the extensive literature on factor analysis and the identification of the factors which together constitute intelligence and learning ability is obvious in this decision.

The attempt to measure three reasonably independent factors with the five tests of the Basic Battery was not very successful. The verbal factor was well covered by the GCT and Reading Tests and also by the Arithmetical Reasoning Test. A separate quantitative factor simply did not appear in the factor analysis of the total battery. A mechanical ability factor appeared to be measured by the battery but less efficiently than was originally hoped. An unplanned factor, not completely clear but possibly measuring ability to handle problems of spatial relations, also appeared in the factor analysis of the battery (see Section 2.4.4).

2.6.2 Usefulness of the Basic Battery Tests

In terms of practical usefulness, the new battery worked well. The tests were of satisfactory reliability. Even against grades assigned in the usual not very systematic manner by a diverse group of instructors in class A Service schools, validity coefficients of .50 to .60 were frequently found for individual tests of the battery.

Table 11[18] shows the mean correlation between each test in Form 1 of the Basic Battery and the final grades achieved by men in 13 types of class A schools. The number of men in each group is also shown. The correlations in the table are those that might be expected if a random sample of recruits were assigned to the schools. These correlations were obtained by correcting the original r's for the homogeneity of the sample.[17] Mean correlations were computed by transmuting each corrected r to z and weighting on the basis of N. Part of the material contained in Table 11, together with a somewhat fuller discussion of the validity data, may be found in reference 21.

TABLE 11. Mean correlations between scores on the Basic Battery tests, Form 1, and final grades in thirteen elementary Service schools (correlations corrected for curtailment in range of talent).

Class A school	Total number of cases	Tests of the Basic Test Battery (1) GCT	(2) READ	(3) ARI	(4) MAT	(5) MK(M)	(6) MK(E)	Best test combination	(7) Multiple corr.	(8) Additive corr.
Aviation ordnancemen	184	.63	.59	.59	.54	.35	.42	GCT & ARI	.66	.66
Basic engineering	1,480	.52	.52	.63	.52	.46	.39	ARI & MK(M)	.66	.63
Diesel (MoMM)	2,160	.42	.35	.36	.26	.43	.46	GCT & MK(E)	.53	.51
Electrical	1,747	.52	.52	.59	.44	.35	.49	ARI & MK(E)	.63	.63
Fire controlman	198	.52	.56	.49	.36	.18	.07	GCT & READ	.58	.58
Gunners mate	1,677	.38	.39	.31	.28	.40	.43	READ & MK(E)	.49	.47
Hospital corps	449	.50	.45	.36	.27	.22	.30	GCT & READ	.51	.51
Machinist mates	755	.33	.27	.44	.48	.48	.45	MAT & MK(E)	.54	.53
Radar operators	1,053	.60	.67	.61	.50	.35	.38	READ & ARI	.70	.70
Signal	984	.49	.43	.44	.25	.13	.13	GCT & ARI	.52	.50
Storekeeper (WR)	678	.44	.47	.59	.37	...	...	READ & ARI	.60	.58
Torpedoman	880	.32	.35	.28	.27	.39	.35	READ & ARI	.44	.43
Yeoman (WR)	738	.62	.59	.63	.38	...	...	GCT & ARI	.67	.67

Table 11 also shows the pair of tests which, in combination, served most efficiently to predict grades in each class A school, the multiple correlation based on those two tests, and their additive correlation. The additive correlations

[redacted]

were obtained by adding the Navy standard scores on the two tests and correlating their sum with school grades. These additive correlations are much easier to compute than are multiple correlations and, for these data, were never more than .03 smaller than the corresponding multiple correlation. When the prediction tests are highly correlated with each other, it may always be true that additive correlations of this type will be as satisfactory as multiple correlations. The method is worth more extensive trial.

The tests of the Basic Battery (Form 1) were also found useful in predicting success in more advanced, class B, schools. The correlations between the tests and the final average course grades achieved by two samples of men enrolled in the class B Gunners Mate School, Washington Navy Yard, are shown in Table 12.[19]

TABLE 12. Correlations between Basic Battery test scores, Form 2, and average grades in class B Gunners Mate School, Washington Navy Yard.

Test and subtest	Correlations with grades Sample 1, $N=84$	Sample 2, $N=163$
General classification test	.51	.55
Sentence completion	.34	.51
Opposites	.33	.42
Analogies	.51	.54
Reading	.47	.63
Arithmetical reasoning	.37	.46
Mechanical aptitude	.58	.51
Block counting	.37	.38
Surface development	.51	.52
Mechanical comprehension	.52	.46
Mechanical knowledge (mechanical)	.53	.44
Tool relationships	.40	.23
Information	.50	.47
Mechanical knowledge (electrical)	.51	.48
Tool relationships	.43	.30
Information	.40	.49
Average of reading and mechanical aptitude tests		.68

Table 13 gives for Form 2 validity information similar to that presented in Table 11 for Form 1. The correlations in Tables 11 and 13 have been corrected for the curtailed range of talent in the sample studied.[17]

During the interval between the introduction of Form 1 and Form 2 of these tests much effort went into the improvement of Service school grades. Particular attention was paid to developing school examinations which placed greater emphasis on ability to perform and less on verbal memory. One important result of this work shows up in comparing the correlations presented in Tables 11 and 13. In general, the correlations of GCT, Reading, and AR with grades in Service schools giving mechanical and electrical training were lower for Form 2 than for Form 1. In contrast, the correlations of MAT and MKT with grades were higher for Form 2 than for Form 1. This result indicates that the mechanical tests (MAT and MKT) served as better predictors of school grades after those grades were made more dependent on ability to perform and less on verbal knowledge.

The data summarized in Tables 11, 12, and 13 indicate very satisfactory validity for the Basic Battery tests. These results, the Navy reports, "fully justify the important place the tests have had in the wartime classification program."[18]

2.7 RECOMMENDATIONS FOR FUTURE WORK

Future work on the development of tests for the initial classification of military personnel can be recommended along two lines.

2.7.1 Better Criteria

Of greatest importance is the development of better criteria for evaluating selection tests. These criteria should be of two types—those from Service schools and those from actual operating performance. The introduction of standardized achievement tests into Service schools (see Applied Psychology Panel, Volume 2, Chapter 17) provides much better school criteria. With improved criteria it becomes possible to determine, much more certainly than one can with poor criteria, which tests are working most efficiently, which should be replaced or revised, how they should be revised, etc. The general problems of securing useful

criteria are discussed in Chapter 17 in Volume 2 of this series.

2.7.2 Special Ability Tests

The success of a military classification program depends upon the accuracy with which many different types of ability can be measured. Ideally, each man should be assigned, within the limits of military requirements, to the work he can do best. Tests which are general in nature, e.g., GCT and MAT, are only partially successful in this respect. They do a good job of picking out the *generally* good men, those who learn easily and who will be successful in any one of a variety of jobs. But which school, or which billet, is best for each man depends upon his special combination of abilities. Measuring these special abilities requires specialized tests. The Speed of Response Test (SOR) (see Chapter 6) for selecting radio code operators is a good example of the type needed.

Good special ability tests will frequently make possible a more equitable distribution of manpower among the various military specialties than can be accomplished with general tests such as the GCT. The SOR, for example, selects the best prospective radio operators. But since scores on the SOR are not highly correlated with scores on the GCT, it does not select the men with highest general ability. It picks good radio operators and leaves the men who are good in other respects, but poor in code ability, for assignment to other types of duty.

The SOR test (see Chapter 6), the tests for radar operators (Chapter 7), stereoscopic visual tests (Chapter 8), night vision tests (Chapter 9), and the Telephone Talker Test (Chapter 10) were all attempts, not always successful attempts, to measure specialized aspects of man's ability—special abilities needed for successful performance in particular military jobs.

Most of these tests differ from the GCT and MAT type of test in that they require some kind of performance other than the writing down of answers to questions. In the SOR, one learns three characters of the radio code. In the Telephone Talker Test one speaks over a sound-powered phone. Tests of this type are called performance tests. Where they duplicate, in miniature, the work situation for which men are being selected, they are called work-sample tests.

Special ability tests are not necessarily performance or work-sample tests, but some of the best ones probably will be. Whether a particular ability can be better measured by a performance test or by a paper-and-pencil test is relatively unimportant. The reliable measurement of a number of special abilities is the important need. Thorough research studies to develop such tests are strongly recommended.

With better criteria to use in validating tests,

TABLE 13. Mean correlations between scores on the Basic Battery tests, Form 2, and final grades in twelve elementary Service schools (correlations corrected for curtailment in range of talent).

School	GCT	Read.	Arith.	MAT	MK(M)	MK(E)	N	No. of school classes
Aviation machinist mate	.440	.315	.305	.485	.645	.560	750	5
Aviation ordnanceman	.535	.390	.350	.285	.355	.490	805	5
Basic engineering	.580	.425	.505	.525	.620	.505	1,176	5
Electrical	.475	.370	.490	.475	.365	.615	1,062	16
Fire controlman	.580	.455	.495	.340	.345	.545	432	5
Gunners mate	.380	.325	.310	.495	.560	.545	809	15
Quartermaster	.535	.485	.460	.320	.205	.270	470	12
Radioman	.270	.255	.270	.165	.195	.235	1,012	16
Storekeeper	.590	.415	.500	.265	.095	.310	541	5
Torpedoman	.290	.255	.245	.330	.540	.390	786	9
Yeoman	.630	.445	.460	.285	.485	.220	857	8
Signalman	.430	.343	.406	.360	.245	.345	864	11

and with better and more diverse measures of specialized abilities, better classification batteries can be constructed. It will then be possible to assign each man to the particular job for which he is best suited, with fewer errors and less difficulty than were experienced in World War II.

It is important to develop better classification batteries. A future war would be even more technically complicated than was the last one; it would require attaining an even closer approximation to the maximum usage of the best abilities of every single person involved, citizen and soldier alike. The development of these improvements takes time. We cannot wait, as we did in World War II, for war to break out before starting to improve military classification tests. They must be developed and experience must be gained in their use, during years of peace.

[redacted]

Chapter 3

THE SELECTION OF OFFICERS

By *Dael Wolfe*[a]

Summary

THE PRINCIPAL UNDERTAKING of the Applied Psychology Panel in the field of officer selection was the development of the U. S. Navy Officer Qualification Test. This test was developed for use by the Offices of Naval Officer Procurement (ONOP's) in appraising the qualifications of men applying for naval commissions. After Form 1 was developed, put into use, and subjected to a detailed statistical analysis, Forms 2 and 3 were developed. They were used in the ONOP's until the end of the war.

Several studies dealt with the selection of officers for special types of duty. Tests developed for radar operators were used in selecting tactical radar officers. The NROTC Selection Test was analyzed, and recommendations for its improvement were made to BuPers. The value of various factors in predicting the success of amphibious flotilla and group commanders was investigated.

A study of combat leadership of infantry officers revealed positive relations of some predictive value with age and with leadership ratings given while at OCS.

3.1 PREPARATION OF U. S. NAVY OFFICER QUALIFICATION TEST

When the Committee on Service Personnel—Selection and Training (which later became the Applied Psychology Panel) was organized in the summer of 1942, the Bureau of Naval Personnel requested a test which could be used to select those officer applicants who had the ability to complete Navy officer training courses satisfactorily. The request specified that the test should be short enough so that it normally would be completed in about 45 minutes. It was required, however, to be a power test, and the candidates were to be allowed approximately an hour and a half to work on the test if they wished. In fulfillment of this request, Project N-106, College Entrance Examination Board, prepared three forms of the U. S. Navy Officer Qualification Test (OQT).

3.1.1 Construction

The content of the Officer Qualification Test (OQT) was of three kinds. Each was included in an effort to measure an ability important in passing an officers' training course. Because one essential qualification for success in officers' training schools is facility in dealing with verbal material, a vocabulary test was included. Because a modern Navy is highly mechanized, a mechanical comprehension test was included in which the problems were presented pictorially. Because quantitative work and training in mathematics are emphasized in the Navy, the third part of the test consisted of arithmetical reasoning problems that involved no formal knowledge of college mathematics. Part One of the experimental form, on vocabulary, contained 60 items; Part Two, on mechanical comprehension, contained 30 items; Part Three, on arithmetical reasoning, contained 25 problems.

The experimental form of the test (called the U. S. Navy Selective Examination, O-1) was tried out at the New York Office of Naval Officer Procurement (ONOP).[1] A sample of 500 papers was obtained for use in item analysis and statistical study designed to aid in the production of the final form of the test.

The odd-even reliabilities of total scores on Form O-1 and of scores on its separate parts are shown in Table 1. The intercorrelations of the subtests and the total test are shown in Table 2.

The total test and the Opposites subtest both

[a] This chapter is based on the work of several of the Applied Psychology Panel projects, chiefly Project N-106.

showed satisfactory reliability. Reliability of the Mechanical Comprehension subtest was too low to be satisfactory. For Arithmetical Reasoning the reliability was borderline. In selecting items for OQT, Form 1, efforts were made to increase the reliability of both of these latter subtests.

TABLE 1. Reliability coefficients for the Selective Examination, O-1.

Opposites	.94
Mechanical Comprehension	.73
Arithmetical Reasoning	.87
Total test	.93

The intercorrelations presented in Table 2 were all satisfactorily low except the one between Mechanical Comprehension and Arithmetical Reasoning, and this correlation turned out to be substantially reduced in Form 1 of the OQT.[7]

TABLE 2. Intercorrelations of parts of the Selective Examination, O-1.

	Mechanical Comprehension	Arithmetical Reasoning	Total test
Opposites	.28	.42	.90
Mechanical Comprehension		.56	.62
Arithmetical Reasoning			.72

Items to be included in Form 1 of the OQT were selected on the basis of detailed item analysis data of the type described in Chapter 14. For each item, both difficulty and correlation with total score on the appropriate subtest were considered.

Fifty items were selected for the Opposites subtest. All had item-test biserial correlations of .45 or higher. The Arithmetical Reasoning subtest was reduced from 25 to 20 items. All those included had item-test correlations of .50 or higher.

Form O-1 included 18 mechanical comprehension subtest items which had been borrowed from the Psychological Corporation. It was decided not to use these in Form 1. Of the 12 original mechanical comprehension items, only 8 had item-test correlations of .35 or higher. These were reused, and 22 new items were constructed for this subtest.

The U. S. Navy Officer Qualification Test, Form 1 is reproduced in the appendix. It consists of 100 items divided as follows: 50 opposites, 30 mechanical comprehension, and 20 arithmetical reasoning. It was designed to be self-administered and to be completed within one hour. The test in this form was turned over to the Navy and used at ONOP's[3] until it was replaced by later forms of the test.

3.1.2 Preparation of Norms

A sample of 4,857 tests administered to men at the various ONOP's during the two-week period ending February 20, 1943, and a similar sample of 5,365 papers from women candidates provided a basis for the preparation of norms for the OQT, Form 1. Separate norms for men and women were furnished to the Bureau of Naval Personnel.[4, 6]

3.1.3 Statistical Analysis

The experimental population used in developing Form 1 of the OQT came entirely from the New York ONOP. Since this regional sample might not fairly represent the whole nationwide population of officer candidates, an examination of this point was called for. Analysis of Form 1 on a new sample of 500 cases from the New York ONOP and a sample of 500 from various ONOP's throughout the United States gave evidence that the difficulty values of the separate items remained satisfactorily constant and that they provided a satisfactory range of item difficulty for the national population of officer candidates.[7]

The biserial correlations between the individual items and the total scores on the appropriate parts of the examination were found to be satisfactorily high for the opposites and arithmetical reasoning portions of the tests. For the mechanical comprehension portion these correlations were somewhat lower on the new sample than they had been on the original sample. In general, however, the correlations

showed only small differences when computed for the United States and the New York samples.

The odd-even reliability of the total test, and of its parts, and the correlations between parts remained essentially the same as those determined in the original development of the test. These data are found in Tables 3 and 4. Comparison of the correlations in Table 3 with those

TABLE 3. Reliabilities of subtests of OQT, Form 1.

	r		$\frac{2r}{1+r}$	
	N.Y.	U.S.	N.Y.	U.S.
Opposites	.87	.85	.93	.92
Mech. Comp.	.47	.45	.64	.62
Arith. Reas.	.76	.75	.86	.86
Total test	.87	.82	.93	.90

in Table 1 shows practically no change in the reliability of opposites and total-test scores. The reliability of arithmetical reasoning scores also remained unchanged. An effort had been made to increase this correlation, but at the same time the number of items was reduced from 25 to 20. The result was a test 20 per cent shorter and equally reliable. Reliability of mechanical comprehension scores dropped from an unsatisfactory .73 to an even more unsatisfactory .62, perhaps because of the necessity of including many new and untried items.

Correlations between parts of the tests were satisfactorily low, lower in fact than the corresponding figures for the experimental version, O-1 (compare Table 4 with Table 2).

In constructing the OQT, which was to be used throughout the United States, the method used was to administer the test in experimental form to a regional sample, analyze it, and choose items for the standard test on the basis of the analysis. The findings summarized above indicated that this method was satisfactory in the development of Form 1 of the OQT. The method is not, however, recommended for general use. Ordinarily it will be safer to use a sample of the total population for which the test is intended.

TABLE 4. Intercorrelations of subtests of OQT, Form 1.

	Opposites		Mech. Comp.		Arith. Reas.		Total test	
	N.Y.	U.S.	N.Y.	U.S.	N.Y.	U.S.	N.Y.	U.S.
Opposites			.23	.16	.42	.36	.91	.90
Mech. Comp.					.47	.38	.55	.50
Arith. Reas.							.70	.66
Mean	27.77	25.73	15.98	16.43	10.23	10.42	53.99	52.49
σ	10.83	10.46	3.73	3.73	4.19	4.03	14.82	13.93

3.1.4 Preparation of OQT, Forms 2 and 3[7]

The OQT, Form 1 was placed in general use early in 1943, without time for careful preliminary standardization and validation. The Bureau of Naval Personnel then asked the Applied Psychology Panel to construct two new forms which could be more carefully standardized and which could replace Form 1. The Bureau desired these two forms to be composed of items with similar difficulty distributions and similar distributions of correlations with total-test score. Items from Form 1 might be reused in Forms 2 and 3. The two new forms were, if possible, to be equated both for men and for women. Forms 2 and 3 were to contain the same number and types of items as Form 1 but were not necessarily to be equivalent to Form 1 in score distribution.

An experimental form of the OQT (Form

O-2), from which items for both Forms 2 and 3 were to be drawn, was developed. This experimental form was administered to about 400 officers at the Naval Training School (Indoctrination) at Dartmouth College, Hanover, New Hampshire, and to about 400 women at the Naval Reserve Midshipmen's School (WR), Smith College, Northampton, Massachusetts. Detailed item analyses were conducted separately for the papers from these two groups.

Items were selected for OQT, Forms 2 and 3 on the basis of the following criteria:[7]

1. High correlation between item and total score on the part of the test containing that item.
2. Item difficulties distributed continuously throughout the range of about .05 to .95 with the greatest concentration near the middle of the distribution.
3. Availability of *pairs* of items having similar difficulty and item-test correlation. One of each pair was included in Form 2, the other in Form 3.

The application of these criteria of item selection made it possible to construct two forms of the tests which were equivalent in difficulty and (odd-even) reliability both for men and for women. The two tests (OQT, Form 2 and OQT, Form 3) are reproduced in the appendix.

3.1.5 Validity[7]

The validity of the OQT, Forms 2 and 3, was studied by determining the correlations of test scores with training school grades at the Naval Training School (Indoctrination), Dartmouth College, Hanover, New Hampshire, and at the Naval Reserve Midshipmen's School (WR), Smith College, Northampton, Massachusetts.

At Dartmouth grades were available for three courses, seamanship, ordnance, and navigation. In addition there was available for each man a rating known as the Officer's Aptitude Rating. This rating was based on judgments made by the instructors as to each man's potential officer qualifications. It was possible to compute reliability coefficients for grades in seamanship and ordnance by correlating the sums of grades received in odd-numbered weeks with the sums of grades received in even-numbered weeks, and then correcting by the Spearman-Brown formula. The reliabilities varied from .75 to .92. The detailed grades necessary for the computation of a reliability coefficient were not available for navigation nor for the Officer's Aptitude Rating.

Correlations between the three forms of the OQT and the grades for two classes at Dartmouth are presented in Table 5. The correlation

TABLE 5. Correlations of OQT with grades in NTSch (Indoctrination), Dartmouth College.

Class 6 $N=360$ Class 7 $N=403$

	OQT Form	Class	Sea.	Ord.	Nav.	Apt.	Final average
Opp.	1	6	.32	.33	.26	.17	.37
	2	7	.37	.25	.35	.02*	.36
	3	7	.38	.27	.34	.03*	.36
Mech. Comp.	1	6	.20	.29	.29	.17	.33
	2	7	.28	.28	.36	.10*	.35
	3	7	.32	.30	.38	.12*	.38
Arith Reas.	1	6	.39	.29	.52	.10*	.49
	2	7	.34	.24	.51	.09*	.41
	3	7	.36	.31	.53	.13	.46
Total test	1	6	.41	.42	.42	.21	.51
	2	7	.45	.34	.51	.07*	.48
	3	7	.47	.37	.50	.09*	.50
Mean		6	3.34	3.35	3.19	3.20	3.26
		7	3.36	3.38	3.12	3.23	3.27
σ		6	.27	.21	.32	.16	.18
		7	.30	.23	.31	.18	.20

*Not significantly different from zero at the 1% level.

of about .50 between total-test score and final average grade was high enough to justify using the OQT as an important element in the selection of officer candidates. The consistency of the correlations with final average grades for the two forms of the test and for different classes (r's of .51, .48, and .50) indicated that the two forms were equivalent and that their usefulness in selecting prospective officers was stable and consistent.

At Smith College each student took either Basic Indoctrination and Communications or Basic Indoctrination and Advanced Indoctrination. These three courses included four subjects each as shown in the following outline:

[redacted]

Basic Indoctrination: (1) Organization, Naval Law, and Naval Correspondence; (2) Personnel; (3) History and Strategy; (4) Ships and Aircraft.

Advanced Indoctrination: (1) Organization and Communications; (2) Personnel; (3) History and Strategy, and Correspondence and Filing; (4) Ships and Aircraft.

Communications: (1) Naval Communications (General); (2) Fundamentals of Radio; (3) Codes and Ciphers; (4) Touch Typewriting.

Estimated reliabilities of the grades were made by correlating grades on two of the subjects in each course with the other two subjects in that course. For the basic indoctrination course these correlations for three companies were .82, .87, and .87. For communications the correlations for two classes were .63 and .69. For advanced indoctrination the correlation for one class was .70. The reliabilities of the individual grades were almost certainly as high as these correlations and probably somewhat higher, since one would expect a subject to correlate higher with itself than with a somewhat different subject.

TABLE 6. Correlations of OQT with grades at Naval Reserve Midshipmen's School (WR), Smith College.

Company 1	Company 2	Company 6
$N = 152$	$N = 160$	$N = 115$

	OQT Form	Co.	Indoc.	Comm.	Adv. Indoc.	Final weighted average
Opposites	2	1	.33	.08*		.20*
		2	.40	.24		.31
		6	.27		.30	.30
	3	1	.34	.10*		.21
		2	.39	.31		.36
		6	.33		.34	.36
Mechanical Comprehension	2	1	.23	.30		.31
		2	.12*	.13*		.14*
		6	.19*		.08*	.15*
	3	1	.19*	.29		.28
		2	.04*	.12*		.11*
		6	.23		.13*	.20*
Arithmetical Reasoning	2	1	.26	.27		.29
		2	.23	.29		.29
		6	.43		.27	.38
	3	1	.32	.31		.34
		2	.29	.27		.30
		6	.48		.36	.45
Total test	2	1	.40	.24		.33
		2	.41	.31		.37
		6	.39		.34	.39
	3	1	.42	.24		.34
		2	.41	.36		.40
		6	.47		.41	.47
Mean		1	3.30	3.16		3.21
		2	3.33	3.19		3.23
		6	3.25		3.27	3.26
σ		1	.24	.21		.20
		2	.25	.21		.21
		6	.25		.21	.22

*Not significantly different from zero at the 1% level.

The correlations of the three forms of the OQT with the grades made by three companies at Smith College are presented in Table 6. The total-test score was found to correlate with the final weighted average grade to the extent of .33 to .40 for those who took the communications course and to the extent of .39 to .47 for those who took advanced indoctrination. The test was therefore shown to work less satisfactorily for women than for men. In view of the differences in validity of the test as applied to the Dartmouth and Smith College populations it was suggested to the Bureau of Personnel that it would be preferable to use Forms 2 and 3 of the OQT for men and to develop a different test for women officer candidates.

3.1.6 Statistical Analysis

After Forms 2 and 3 of the OQT had been put into use, answer sheets for a sample group of candidates were collected from the ONOP's. Item difficulties, biserial correlations of each item with the score on the appropriate subtest, and reliabilities and intercorrelations of the subtests were determined.[10]

On the basis of this statistical information, the following recommendations were made to the Bureau of Naval Personnel:

1. Whenever practicable, experimental tests should be tried out on a sample exactly similar to the group for whom the regular form of the test was intended. It would have been better to have tried out the experimental edition of Forms 2 and 3 on a group of applicants rather than on a group of already successful

candidates. The men in the school group differed slightly in average ability from the officer candidates at the ONOP's and showed a smaller range of ability.

2. An attempt to increase the reliability of the mechanical comprehension subtest of the OQT was recommended. Its reliability of .74 to .78 was too low to be acceptable.

3. The three subtests were sufficiently independent to justify reporting separate scores rather than one total score.

Forms 2 and 3 of the OQT continued in use at the ONOP's until the end of the war.

3.2 OTHER STUDIES OF OFFICER SELECTION

3.2.1 The Prediction of Combat Leadership

In fulfillment of a request from the Office of the Adjutant General, War Department, the Applied Psychology Panel conducted a preliminary study of combat leadership. Two classes of infantry officer candidates and two classes of field artillery officer candidates were studied. For each of the four classes, correlations were computed between graduation or failure to graduate and a number of other variables which might be used to predict officer success. In addition, combat ratings of 176 infantry company officers were compared with several types of officer candidate school (OCS) records.[9]

Prediction of Success in Infantry OCS

Information was obtained on 40 items concerning the men in two classes at Fort Benning. Many of these items were found to be completely unrelated to success or failure in OCS. Those which appeared to be of greatest possible usefulness are given in Table 7 together with their correlations with OCS success or failure. Some of the correlations presented in this table are at variance with common belief, for example, the finding that hobbies, classified as vigorous versus sedentary, showed a zero correlation with the criterion of being commissioned or not being commissioned.

TABLE 7. Correlations of seven items obtained before or upon entrance to infantry OCS against the criterion of being commissioned or not being commissioned.

Criterion vs following items*	Class 304 $N =$ about 200	Class 311 $N =$ about 200
1. Average first and second AGCT scores	.45	.46
2. Amount of schooling	.39	.07
3. Age at entrance to OCS	—.12	—.04
4. Father's occupation (leadership vs nonleadership)	.12	.32
5. Athletic participation	—.08	—.10
6. Hobbies (hobbies classified as vigorous and athletic vs mild and sedentary)	.00	.02
7. Leadership as indicated by former civilian positions	.20	.40

* Correlations 1, 2, 3 are biserial; 4, 5, 6, 7 are tetrachoric.

At this OCS, both platoon leaders and fellow officer candidates made periodic ratings or rankings of the leadership qualities of the members of their platoon. These ratings were considered in deciding which candidates should be commissioned. Some of the correlations of the ratings with the criterion are presented in Table 8.

TABLE 8. Biserial correlations of leadership ratings against criterion of being commissioned or not being commissioned in infantry OCS.

Criterion vs following items	Class 304	Class 311
Platoon leaders' ratings at the end of 5 weeks	.36 ($N = 201$)	.84 ($N = 204$)*
Platoon leaders' ratings at the end of 10 weeks	.70 ($N = 143$)	.78 ($N = 143$)
Platoon leaders' ratings at the end of 15 weeks	.85 ($N = 145$)	.84 ($N = 126$)
Final company ratings	.87 ($N = 145$)	.90 ($N = 126$)

* This high r was apparently due to the large number of ROTC candidates in Class 311. These men had received considerable military training and their leadership qualities were apparently evaluated quickly by their platoon leaders.

Table 9 shows the intercorrelations among selected items of information obtained.

None of the items obtainable before or upon entrance to OCS was highly predictive of later

[redacted]

performance in school. The highest correlation was with the Army General Classification Test score. The next highest was with leadership as indicated by former civilian positions. This measure looks hopeful for selection purposes. Attempts to measure the leadership actually shown in civilian positions would be well worth while. The information gained might materially improve the selection of officer candidates.

TABLE 9. Intercorrelations among selected items at infantry OCS.

	Class 304	Class 311
First AGCT and second AGCT	.64 ($N = 213$)	.72 ($N = 212$)
Platoon leaders' ratings for leadership, end of 5th week, and average AGCT	.21 ($N = 200$)	.20 ($N = 202$)
Platoon leaders' ratings for leadership, end of 5th week, and highest military rank attained in college ROTC	—.16 ($N = 37$)	.15 ($N = 119$)
Platoon leaders' ratings for leadership, end of 5th week, and end of 15th week	.65 ($N = 99$)	.49 ($N = 114$)

PREDICTION OF SUCCESS IN FIELD ARTILLERY OCS

Data from the Field Artillery Officer Candidate School at Fort Sill are given in Tables 10 and 11.

Background information available upon entrance to OCS was generally more useful in predicting success of field artillery officer candidates than it was for infantry officer candidates. Previous education and grades on mathematics examination scores both showed useful relationships (correlations of .33 to .48) with success in field artillery OCS.

The differences in the correlations found for classes 80 and 82, which sometimes were very large, appeared to be due to differences in the men enrolled in the two classes. Class 80 consisted of men from all branches of Service and included a number who had been in the Army for a considerable period. Class 82 consisted almost entirely of ROTC candidates. The differences in correlations for the two classes

TABLE 10. Biserial correlations of nine items obtained before or upon entrance to field artillery OCS against the criterion of being commissioned or not being commissioned.

Criterion vs following items	OC 80 $N =$ about 200	OC 82 $N =$ about 200
Army General Classification Test	.29	.58
Education (number years in school beyond elementary level)	.33	.36
Age	—.22	—.67
Participation in athletics	—.20	.16
Civilian occupation (rating from Barr Scale)	.17	.32
Length of military service	—.52	.74
Mathematical test (given first week of course)	.48	.40
Mathematical courses (number previously taken in school)	.46	.00
Administration and military law (test given on arrival)	.39	.12

TABLE 11. Biserial correlations of ratings of leadership against the criterion of being commissioned or not being commissioned in field artillery OCS.

Criterion vs following items	OC 80 $N =$ about 200	OC 82 $N =$ about 200
Tactical officers' ratings at end of 5th week (leadership, courage, personal appearance)	.80	.74
Tactical officers' final ratings before graduation (leadership, courage, personal appearance)	.90	.80
Gunnery officers' ratings. First rating during 5 weeks' course (ability shown in firing problems)	.52	.86
Gunnery officers' ratings. Second rating during 5 weeks' course	.88	.76
Ratings by classmates at end of 5th week	.54	.70
Ratings by classmates just before graduation	.96	.90

made prediction hazardous, but, in general, it appeared probable that success in field artillery OCS could be predicted more accurately from data available before entrance than was the case for infantry OCS.

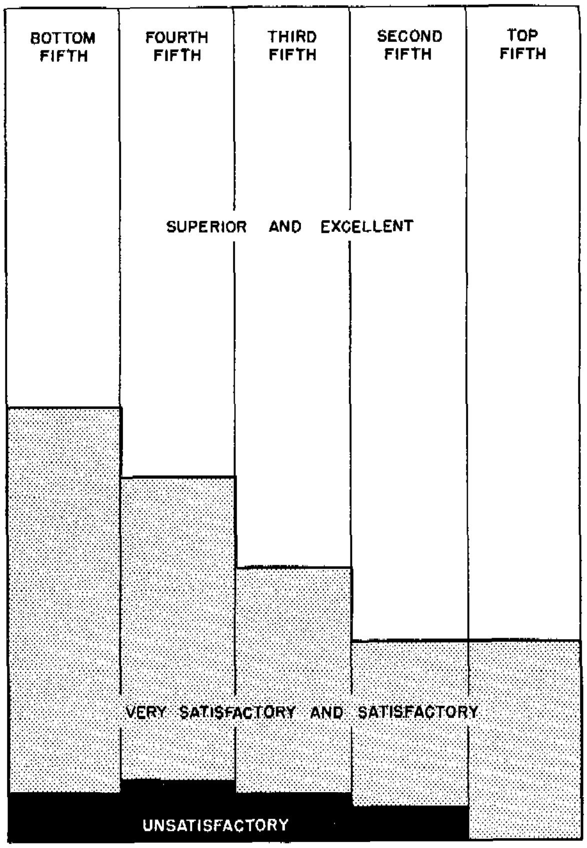


FIGURE 1. Agreement between combat ratings and ratings of leadership qualities shown in infantry OCS.

The Relations between Success in Infantry OCS and Combat Ratings

Ratings were obtained on the combat efficiency of 176 company officers all of whom had graduated from the infantry OCS at Fort Benning. These ratings were made on a five-point basis by the regimental commanders or their executive officers and by the battalion commanders. The ratings were distributed as follows:

Rating	Number	Per cent
Superior	23	13.1
Excellent	87	49.4
Very Satisfactory	41	23.3
Satisfactory	17	9.7
Unsatisfactory	8	4.5

The agreement between combat ratings and the leadership ratings assigned while in school at Fort Benning can be seen in Figure 1. All the men who received school ratings in the first (top) fifth for leadership received combat ratings of "Satisfactory" or better. The percentages of combat ratings of "Superior" and "Excellent" became steadily smaller for the groups which received lower leadership ratings in school.

There was no significant relation between combat success and AGCT score. But the group was a select one to begin with, the average AGCT score being 124.6.

There was a fair relation between age and combat rating. The superior and excellent officers on the whole came from the age group 22 to 28. Proportionately fewer men below 22 or above 28 were so rated.

After these men graduated, the academic requirements were changed. The effect which the changed standards would have had on these 176 officers is shown below.

	Per cent who would not have graduated under changed requirements	Number who would not have graduated under changed requirements
Superior group	4.3	1
Excellent group	10.3	9
Very Satisfactory group	24.3	10
Satisfactory group	29.4	5
Unsatisfactory group	50.0	4

The trend of percentages shown by these figures is very encouraging. The actual figures on which they are based, however, show that the gain of graduating four fewer officers who later received unsatisfactory combat ratings would have been made at the expense of rejecting 20 men who later received combat ratings of "Very Satisfactory," "Excellent," or "Superior." When officer candidates are plentiful this expense is justified.

Conclusion Regarding Combat Leadership

The results of this brief study of combat leadership were encouraging enough to warrant further investigations in this field. Certainly the importance of accurate prediction of successful combat leadership is great enough to

justify a good deal of effort directed toward improving the methods used to select company officers. Of first importance in such a study will be the task of obtaining careful, detailed, and reliable measures of success or failure on the part of a large number of officers engaged in a variety of military duties. These measures are necessary as a criterion to use in evaluating possible selection tests or measures.

3.2.2 NROTC Selection Test[2]

The Bureau of Naval Personnel requested the Applied Psychology Panel to make an analysis of the NROTC Selective Examination and to suggest revision of that test. The test consisted of four parts: 100 multiplication problems; 50 block counting items; 50 vocabulary items; and 50 arithmetical reasoning items.

The study of the NROTC Selective Examination included an item analysis of a sample of 500 papers, a study of the reliability and intercorrelations of the parts of the test, and a study of the validity of the test. Grades in deck officers school ($N = 216$) and engineering school ($N = 193$) were used as criteria. This investigation led to the conclusion that the total score could be made somewhat better for predicting grades by omitting the multiplication section and by simply summing the scores on the other three parts instead of using the complicated derived-score system previously in use. For two groups of students, correlations between total derived scores on the test and final grades were .49 and .51. Using unit weights for the three tests (Block Counting, Vocabulary, and Arithmetical Reasoning) gave multiple validity coefficients of .50 and .55 for the same two groups. Omitting the Block Counting test and weighting the other two equally reduced these correlations to .45 and .49. It therefore seemed probable that the block counting section could also be eliminated and the vocabulary and arithmetical reasoning sections improved sufficiently to give validity coefficients as good as those found for all three tests. Detailed suggestions for revising these two subtests were made to the Bureau of Naval Personnel.[2]

3.2.3 Selection of Flotilla and Group Commanders[11]

Project N-117 of the Applied Psychology Panel conducted a brief investigation of the information which might be useful in selecting flotilla and group commanders in the Amphibious Training Command, U. S. Atlantic Fleet. The study was made at the request of that Command.

Nine possible predictors were studied in relation to fitness report data (Items 8, 10, and 13 of the Report on the Fitness of Officers, NavPers 310) for a group of 76 officers, 42 of whom were graduates of the Naval Academy and 34 of whom were not.

The predictors included in the study were OQT, Form 3; MKE, Form 2 and MKM, Form 2 (see Chapter 2); Officer Personnel Inventory (see Chapter 4); age, rank, months of sea duty, whether an academy graduate or not, and interviewer's overall evaluation.

The interviewer's evaluation showed higher correlations (.40, .36, and .45) with the three fitness report items than did any of the other predictors. The small number of cases included in the study and the fact that both the interviewer's opinion and the fitness reports were, in part, based upon the same factors and the same kind of general opinion—both require that these results be interpreted with caution.

3.2.4 Selection of Tactical Radar Officers

Two projects of the Applied Psychology Panel contributed significantly to the development of tests and procedures used by the Navy to select officers for duty in combat information centers. These tests are described in Chapter 7, "Selecting Radar Operators."

[redacted]

Chapter 4

ELIMINATING THE EMOTIONALLY UNFIT

By *Dael Wolfle*[a]

Summary

THE PERSONAL INVENTORY was developed in order to have available a device for making quick identification of emotionally unstable men who are likely to break down under the stress of hazardous duty.

The Personal Inventory consists of paired statements. From each pair, the man being tested selects the one which best fits him. Each pair includes a symptomatic and a nonsymptomatic (normal) statement. The number of symptomatic statements checked constitutes a man's score.

With many different Service groups it was demonstrated that men making high scores were much more likely to be judged unfit for service, when examined by a psychiatrist, than were men making low scores. The Personal Inventory was found to discriminate better than other tests and methods tried out between acceptable men and those given psychiatric discharges.

The Personal Inventory was used as a preliminary screening device to aid the psychiatrist by identifying the men most likely to be classed as emotionally unfit. It was thus possible for the psychiatrist to interview only a fraction of the men, but to find, in that fraction, most of those who would have been given psychiatric discharges if all men had been interviewed.

4.1 INTRODUCTION

In selecting men for special types of duty, such as officer candidate schools, submarines, parachute troops, etc., physical examinations weed out the obviously physically unfit, and aptitude tests weed out most of those too inapt to learn their special duties. A group is left who are disqualified because they are emotionally unstable and likely to break down under the stress of hazardous duty. Physical examinations and aptitude tests are not designed to pick out this type of person, and they do not do it well.

Brief psychiatric examinations can sometimes be used to identify the emotionally unstable or psychiatric cases. But the limited number of qualified military psychiatrists and the large number of men to be examined combined, during World War II, to make it impossible to give thorough psychiatric examinations to all men for whom such examinations would have been desirable.

Civilian experience prior to the war led to the belief that it would be valuable to develop a test for preliminary screening of psychiatrically undesirable men in the military Services. It was agreed that an inventory should be developed which would emphasize biographical items. A contract to develop such an inventory was written with Brown University. The project started under Section D-2 of NDRC, continued under the National Research Council Committee on Service Personnel—Selection and Training, and was completed under the Applied Psychology Panel as Project N-113.

The result of this work was the Personal Inventory. In many Service situations it proved to be a highly useful device for the preliminary screening of large groups of men. Those showing evidence of emotional disturbance or instability could be interviewed by the psychiatrist while the others could be passed without interview. Where regulations required interviewing every man, more time could be spent with those whose scores indicated considerable likelihood of emotional instability.

4.2 THE PERSONAL INVENTORY

Four forms of the Personal Inventory were developed. Three were designed for use with enlisted men and one for use with officers.

[a] This chapter is based primarily upon the work of NDRC Project N-113.

4.2.1 Enlisted Men's Personal Inventory, Long Form[1]

DEVELOPMENT

The original form of the Personal Inventory (PI) for enlisted men contained 145 items, most of which were in a *forced-choice* form. In answering each item one must choose between two statements. One alternative in each item presents a symptomatic attitude or type of behavior; the other a nonsymptomatic attitude or type of behavior of approximately equal acceptability. For example:

I have felt bad more from head colds ( )	I have felt bad more from dizziness ( )

For several reasons the items were phrased in this *forced-choice* form rather than in the more familiar *Yes-No* or *Yes-No-Doubtful* form. The most important reason was the belief that this form would elicit more honest answers than a *Yes-No* question on the same symptom, e.g.:

I frequently feel dizzy Yes No

In addition, statistical treatment of the answers for the form used was relatively easy, and it was believed less subject to a "halo" effect than the *Yes-No* form of item. One disadvantage of the form was the fact that many men objected to being forced to choose between two statements neither of which applied particularly well to them. In spite of this objection, the method worked very effectively.

The items were constructed on the basis of a careful study of the case histories of a group of normal enlisted men and another group of psychiatric cases. The case histories were obtained from an OSRD report entitled "Analysis of 100 Psychiatric Cases from the Chelsea Naval Hospital." All case records were examined for biographical items which appeared to differentiate the normal from the psychiatric histories. The term "biographical" was construed broadly to include educational, occupational, social, and attitudinal factors and any other symptoms which occurred significantly more frequently in one group than in the other. Analysis of the case histories was supplemented by conferences with naval psychiatrists and other medical officers and by reference to naval and military psychiatric literature.

The 300 items originally constructed were sifted to give appropriate emphasis to each area to be covered in the final inventory. Those finally selected were then further refined to make them understandable and relatively inoffensive to the subjects and to make the two alternatives in each item approximately equally desirable. In most items, however, there was a slightly greater apparent desirability for the choice usually made by normal individuals, the nonsymptomatic alternative.

The inventory was constructed so that the answers could be recorded on a separate answer sheet which could be scored either by hand or by machine.

The inventory was printed in two formats. In Format A the answers were recorded through holes in the test booklet. The scoring key collected together the answers to related items. This feature made it possible to secure subscores on each of 20 separate clusters of items, for example, hypochondriasis, sociability, or mood swing. Format B did not have this clustering principle or the mechanical elaborateness which it required. In Format B only a single score was obtained. The two formats were shown experimentally to be equivalent.[8] In practice, Format B soon became widely used and Format A practically forgotten. Since the items on the two formats were identical, further discussion of the long form will not distinguish between Formats A and B. In practically all cases Format B was the one used. That form is reproduced in the appendix.

A later study[19] from the Medical Research Department, U. S. Submarine Base, New London, Connecticut, supplied statistical support for the general preference for Format B over Format A. This study presented the intercorrelations of 58 of the 60 scored items. Of a total of 1,653 (tetrachoric) correlations, 1,220 fell within the range $-.10$ to $+.19$. A detailed factor analysis of the data was not made, but

TABLE 1. Item analysis of the Personal Inventory, long form.

		N answering item		*N* answering given alternative		Percentage answering given alternative		
Item No.	Choice	Nor-mals	Dis-charges	Nor-mals	Dis-charges	Nor-mals	Dis-charges	Critical ratio
1	R	1,001	384	190	113	19.0	29.4	4.3
2	R	994	380	796	306	80.1	80.5	0.4
3	R	980	360	814	302	83.1	83.9	0.4
4*	R	956	362	130	121	13.6	33.4	7.0*(*S*)
5*	R	994	381	179	116	18.0	30.4	4.6
6	R	1,000	379	81	56	8.1	14.8	3.5
7	R	995	380	358	164	36.0	43.2	2.4
8*	R	1,003	385	37	41	3.7	10.6	4.1
9	R	1,004	385	93	38	9.3	9.9	0.6
10	R	994	384	61	26	6.1	6.8	0.7
11*	R	998	384	139	120	13.9	31.3	6.5*
12	L	996	383	161	128	16.2	33.4	6.3*
13*	R	1,000	384	67	50	6.7	13.0	3.2
14	L	1,001	385	250	118	25.0	30.6	2.2
15	L	1,001	385	190	120	19.0	31.2	4.5
16*	R	994	384	309	192	31.1	50.0	5.5*
17	R	1,001	381	187	105	18.7	27.6	3.5
18	R	990	380	120	70	12.1	18.4	2.7
19	L	999	384	107	88	10.7	22.9	5.2
20	R	1,000	384	119	72	11.9	18.8	3.1
21	R	1,001	383	133	103	13.3	26.9	5.5
22	R	996	384	503	231	50.5	60.2	3.0
23	L	994	382	184	108	18.5	28.3	3.5
24	L	992	382	262	118	26.4	30.9	1.8
25	L	1,001	385	93	77	9.3	20.0	5.0
26	L	1,003	385	356	176	35.5	45.7	3.4
27	L	996	384	131	93	13.2	24.2	3.7
28	L	1,001	384	25	26	2.5	6.8	9.1*
29	L	1,003	385	269	149	26.8	38.7	4.2
30*	R	997	383	274	193	27.5	50.4	6.5*
31	R	999	381	191	161	19.1	42.3	9.1*(*S*)
32*	L	999	379	331	223	33.1	58.8	8.9*
33	R	998	384	694	290	69.5	75.5	2.6
34	R	988	376	60	57	6.1	15.2	4.6
35*	L	998	374	530	275	53.1	73.5	7.5*
36*	R	997	367	34	143	3.4	39.0	14.1*(*S*)
37	R	1,000	378	461	214	46.1	56.6	2.8
38	L	998	385	407	224	40.8	58.2	5.7*
39*	R	1,003	385	53	98	5.3	25.5	8.7*(*S*)
40	L	1,004	385	466	182	46.4	47.3	0.3
41	L	1,002	385	424	219	42.3	56.9	5.1
42*	R	1,003	383	119	158	11.9	41.3	10.8*(*S*)
43	L	1,000	381	700	318	70.0	83.5	3.0
44	L	999	383	214	103	21.4	26.9	2.5
45	L	1,002	383	103	67	10.3	17.5	3.7
46	R	1,002	383	873	342	87.1	89.3	1.0
47*	L	999	385	12	157	1.2	40.8	15.9*(*S*)
48*	R	994	380	93	53	9.4	13.9	2.4
49	R	969	370	585	231	60.4	62.4	0.3
50*	R	986	380	115	98	11.7	25.8	5.6*
51*	L	1,001	384	144	211	14.4	54.9	15.0*(*S*)
52*	L	1,000	382	10	51	1.0	13.4	6.9*
53	R	1,002	385	275	194	27.4	50.4	7.7*
54	L	1,002	383	533	269	53.2	70.2	6.1*
55	L	1,000	384	517	257	51.7	66.9	5.2
56	L	998	380	337	175	33.8	46.1	3.5
57	L	994	385	106	58	10.7	15.1	1.8
58	R	1,002	385	317	133	31.6	34.5	1.0
59*	L	1,001	382	143	109	14.3	28.5	5.9*
60	L	999	384	149	71	14.9	18.5	1.8
61*	L	999	380	214	171	21.4	45.0	7.1*(*S*)
62*	R	992	377	477	228	48.1	60.5	4.0
63*	L	972	366	196	197	20.2	53.8	9.8*(*S*)
64	R	1,000	379	695	269	69.5	71.0	0.7
65*	R	999	381	119	142	11.9	37.3	5.6*(*S*)
66*	R	998	377	41	109	4.1	28.9	10.5*(*S*)
67	R	998	384	294	147	29.5	38.3	3.2
68	L	993	383	262	135	26.4	35.2	2.9
69*	L	995	381	96	152	9.6	39.9	11.5*(*S*)
70*	L	1,002	383	217	165	21.7	43.1	7.5*
71	R	997	381	182	99	18.3	26.0	3.2
72	R	994	382	104	87	10.5	22.8	5.0
73	L	998	382	341	168	34.2	44.0	3.4
74*	L	997	381	485	263	48.6	69.0	6.9*
75*	L	1,000	384	180	152	18.0	39.6	7.9*(*S*)
76	L	1,002	385	211	112	21.1	29.1	6.2*
77	L	1,004	384	211	100	21.0	26.0	2.0
78*	L	994	380	247	199	24.8	52.4	7.9*
79*	L	990	374	181	118	18.3	31.6	5.2
80*	L	998	383	65	108	6.5	28.2	8.6*
81	L	1,003	385	456	194	45.5	50.4	1.4
82*	L	1,001	385	102	195	10.2	50.6	15.4*(*S*)
83	R	1,002	378	823	326	82.1	86.2	1.8
84	R	999	383	261	202	26.1	52.7	7.9*
85	L	1,001	385	218	92	21.8	23.9	0.8
86	L	1,002	384	461	198	46.0	51.6	2.0
87	R	1,000	384	565	266	56.5	69.3	4.2
88	L	994	382	526	214	52.9	56.0	1.0
89	L	1,001	385	61	49	6.1	12.7	3.7
90*	L	991	378	45	48	4.5	12.7	4.4
91*	R	999	382	9	71	0.9	18.6	8.9*
92	R	1,001	382	142	88	14.2	23.0	3.7
93	L	999	383	118	70	11.8	18.3	2.7
94*	L	1,003	385	127	144	12.7	37.4	9.5*
95	L	1,002	383	199	99	19.9	25.8	2.3
96*	L	1,001	384	22	94	2.2	24.5	9.8*(*S*)
97*	L	996	384	486	241	48.8	62.8	4.7
98*	R	1,001	383	163	163	16.3	42.6	9.8*
99	L	1,002	381	371	151	37.0	39.6	1.0
100*	L	1,003	384	158	117	15.8	30.5	5.4
101	L	999	385	289	165	28.9	42.9	4.9
102*	R	1,003	384	152	139	15.2	36.2	8.0*
103*	L	1,003	384	534	298	53.2	77.6	9.5*(*S*)
104	R	1,004	383	183	126	18.2	32.9	5.6*
105*	R	998	384	402	242	40.3	63.0	7.7*
106	R	1,000	384	260	145	26.0	37.8	4.2
107*	L	1,002	385	50	133	5.0	34.5	12.0*(*S*)
108*	L	997	382	343	242	34.4	63.4	9.9*
109*	R	999	381	16	60	1.6	15.7	7.2*
110*	R	1,000	384	26	100	2.6	26.0	10.2*(*S*)
111	R	996	382	338	130	33.9	34.0	0.0
112*	R	986	378	30	65	3.0	17.2	7.1*
113*	L	1,002	384	45	92	4.5	24.0	8.2*
114	R	996	379	98	65	9.8	17.2	3.3
115*	L	1,002	383	39	65	3.9	17.0	6.5*
116	R	1,001	385	56	30	5.6	7.8	1.2
117	L	988	382	529	243	53.5	63.6	3.4
118	L	934	370	112	54	12.0	14.6	1.4
119	L	1,000	382	299	189	29.9	49.5	5.6*
120	L	1,003	384	61	40	6.1	10.4	2.4
121*	L	996	382	209	188	21.0	49.2	8.3*
122*	L	1,003	383	19	76	1.9	19.8	8.8*
123*	R	996	382	26	120	2.6	31.4	11.9*(*S*)
124	L	998	380	633	266	63.4	70.0	2.5
125	R	991	382	20	12	2.0	3.1	1.0
126*	L	992	379	89	83	9.0	21.9	5.7*
127	L	1,000	383	244	123	24.4	32.1	2.9
128	L	1,001	385	157	87	15.7	22.6	2.9
129*	L	1,001	385	109	135	10.9	35.1	9.2*
130	L	996	384	234	137	23.5	35.7	4.3
131	L	1,000	381	122	94	12.2	24.7	5.4
132	L	1,000	384	156	98	15.6	25.5	4.0
133	L	995	378	98	48	9.8	12.7	1.5
134	R	998	381	26	11	2.6	2.9	0.0
135	L	998	383	152	72	15.2	18.8	1.8
136	L	997	385	228	160	22.9	41.6	6.7*
137*	L	998	384	95	118	9.5	30.7	8.2*(*S*)
138*	R	997	385	364	243	36.5	63.1	9.3*
139*	L	1,003	383	22	54	2.2	14.1	6.5*
140*	L	1,001	384	334	195	33.4	50.8	5.2
141*	R	998	384	14	52	1.4	13.5	7.1*
142*	L	1,001	385	50	75	5.0	19.5	7.1*
143*	L	1,001	383	10	31	1.0	8.1	4.9
144	R	999	384	74	29	7.4	7.6	0.6
145	R	1,000	385	139	79	13.9	20.5	3.0

Note—The item numbers are shown in column 1. The 60 items marked with an asterisk are the 60 most discriminative items found in the first item analysis.[1] The 60 items marked with an asterisk in the last column are the 60 most discriminative items found in the second item analysis.[6] The 20 critical ratios in the final column marked with the letter (*S*) indicate the 20 items which later became the short form of the Personal Inventory.[7] The choice indicated in the second column is the choice marked more frequently by the discharge group than by the normal group.

[redacted]

inspection seemed to indicate the presence of two factors. There was no inspectional support for the use of the twenty separate cluster scores obtained with Format A. Correlations between answers to 37 PI items and Otis Intelligence Test scores are also contained in reference 19.

SCORING KEY

A scoring key[1] was devised by administering the long form to a group of 1,210 newly enlisted men who had been passed upon favorably in a psychiatric interview, and to 105 men who had been passed upon by a discharge board and were being discharged for psychiatric reasons. These men were tested while in the psychiatric ward. Both groups were obtained at the U. S. Naval Training Station, Newport, Rhode Island.

Papers were excluded for all men who omitted more than five items and for those who had at some time been under psychiatric observation but had later been returned to their training companies for trial duty. These exclusions left a group of 1,004 normals and a group of 84 psychiatric discharges.

An item analysis of the answers of these two groups to each of the 145 items was made. Differentiation of each item between the two groups was measured in terms of the critical ratio. Sixty items were found to have critical ratios of 2.7 or greater. The scoring key for the long form of the PI is based on these 60 items. A man is given one point for each of the 60 items to which his answer is the same as that characteristically given by the discharge group. Low scores are, therefore, "good" scores.

A second analysis[5] of the same type was made on a new group of 385 psychiatric discharges from Newport Naval Training Station and the same group of 1,004 Newport normals. Forty-nine of the 60 items having critical ratios of 2.7 or greater in the first analysis were among the 50 most discriminative items in the second analysis. The detailed results of this second item analysis are given in Table 1. Comparison of the ability of the PI to distinguish between normal and psychiatric groups showed only insignificant differences between the two scoring keys. The original scoring key was therefore retained.

The distribution of critical ratios of the 60 scored items is given in Table 2. The median critical ratio was over 4.0, and all were greater than 2.7.

TABLE 2. Critical ratios of items in the Personal Inventory scoring key.

C. R. (diff. /$\sigma_{diff.}$)	Number of items
2.7–2.9	4
3.0–3.9	24
4.0–4.9	17
5.0–5.9	10
6.0–6.9	3
7.0–7.9	2
	60

RELIABILITY

The reliability of scores on the long form was determined on four groups of subjects. The data are given in Table 3.

TABLE 3. Reliability coefficients (odd-even, Spearman-Brown corrected) for the long form of the Personal Inventory.

Group	Mean	σ	r_t
508 Newport recruits	9.4	4.6	.69
610 amphibious forces men	14.4	8.1	.85*
188 Newport psychiatric discharges	25.8	11.1	.91
124 Newport psychiatric discharges	19.8	10.9	.92

* This reliability coefficient was based on a first half–second half rather than an odd-even correlation.

VALIDITY

To test the validity of the PI it was administered to several new groups and scored by the previously devised stencil.[1] These new groups were obtained at the U. S. Naval Training Station, Newport, Rhode Island, the U. S. Naval Hospital, Chelsea, Massachusetts, and the U. S. Submarine Base, New London, Connecticut. The group at Newport Naval Training Station consisted of 508 inductees who had passed through psychiatric screening and were considered normal; 116 men who had

been under observation in the psychiatric ward but who were subsequently returned to their companies for trial duty; and 124 men who were being discharged for psychiatric reasons.

The answer sheets for all three groups were scored by means of the previously devised stencil. The number and cumulative percentage of each of the three groups falling at each score is given in Table 4.

TABLE 4. Number and cumulative percentage at each score on Personal Inventory for 508 normals, 124 discharges, and 116 back-to-duty cases at Newport Naval Training Station.

508 Newport normals			124 Newport discharges			116 Newport back-to-duty		
Score	*N*	Cumulative per cent	Score	*N*	Cumulative per cent	Score	*N*	Cumulative per cent
0	1	0.19	0	0	....	0	0	....
1	5	1.17	1	0	....	1	2	1.72
2	21	5.30	2	2	1.61	2	4	5.17
3	17	8.64	3	0	1.61	3	5	9.48
4	34	15.35	4	2	3.22	4	6	14.66
5	33	21.82	5	3	5.64	5	8	21.56
6	32	28.11	6	5	9.68	6	11	31.06
7	40	36.08	7	2	11.29	7	10	39.68
8	38	43.56	8	4	14.52	8	15	52.58
9	47	52.81	9	5	18.56	9	10	61.20
10	51	62.84	10	6	23.40	10	11	70.70
11	40	70.71	11	9	30.65	11	5	75.01
12	35	77.69	12	1	31.46	12	10	83.63
13	22	81.92	13	9	38.71	13	2	85.35
14	25	86.84	14	2	40.32	14	1	86.21
15	19	90.58	15	4	43.55	15	4	89.66
16	17	93.82	16	4	46.78	16	0	89.66
17	11	96.08	17	2	48.39	17	3	92.25
18	7	97.45	18	4	51.62	18	0	92.25
19	3	98.04	19	2	53.23	19	2	93.97
20	3	98.63	20	4	56.46	20	2	95.69
21	0	98.63	21	3	58.88	21	1	96.55
22	3	99.22	22	2	60.49	22	0	96.55
23	1	99.41	23	6	65.33	23	1	97.41
24	1	99.60	24	5	69.37	24	1	98.27
25	0	99.60	25	3	71.79	25	0	98.27
26	1	99.89	26	1	72.60	26	1	99.13
27	0	99.89	27	4	75.83	27	0	99.13
28	0	99.89	28	0	75.83	28	0	99.13
29	0	99.89	29	4	79.06	29	1	100.00
30	0	99.89	30	1	79.87			
31	0	99.89	31	2	81.48			
32	0	99.89	32	5	85.52			
33	0	99.89	33	0	85.52			
34	1	100.00	34	2	87.13			
			35	0	87.13			
			36	2	88.74			
			37	6	93.58			
			38	1	94.39			
			39	1	95.20			
			40	1	96.01			
			41	2	97.62			
			42	1	98.43			
			43	0	98.43			
			44	0	98.43			
			45	1	99.20			
			52	1	100.00			

It may be seen from Table 4 that cutoff scores can be selected which include most of the discharges and only a small percentage of the normals. For example, 3.9 per cent of the normals made scores of 18 or higher, while 51.6 per cent of the discharges made such scores. The difference between the percentages of discharges and normals identified by a cutting score of 18 was over 12 times its standard error.[1]

Similar analyses of the answer sheets from a smaller group of men tested at the U. S. Naval Hospital, Chelsea, Massachusetts, showed essentially the same ability to distinguish between normal and psychiatric cases. A critical score of 18 identified 60 per cent of the psychiatric cases while including 3.9 per cent of the normals. Further evidence supporting the value of the inventory came from analysis of the New London data.

A comparison of average scores made by different Navy groups to whom the long form of the PI was given supplied additional information on the validity of the Inventory. These data are presented in Table 5.[6]

TABLE 5. Personal Inventory scores for various Navy groups.

		Mean Personal Inv. score	Standard deviation
276	Commissioned officers in submarine school	6.4	3.9
646	Coast Guard Reserve Cadets	7.4	4.6
189	Psychiatrically approved submarine school men	8.3	4.3
609	Newport normal recruits subsequently receiving ratings	8.3	4.2
1,512	Newport normal recruits	9.3	4.6
856	Newport normal recruits not receiving ratings within a year	10.0	4.8
444	USS *New Jersey* men	11.1	5.7
74	Psychiatrically disapproved submarine school men	14.6	8.4
2,301	Amphibious Forces men	15.2	8.0
301	Newport psychiatric discharges	21.6	11.0

The differentiating ability of the Inventory is shown graphically in Figure 1. The separate points through which the two curves are drawn

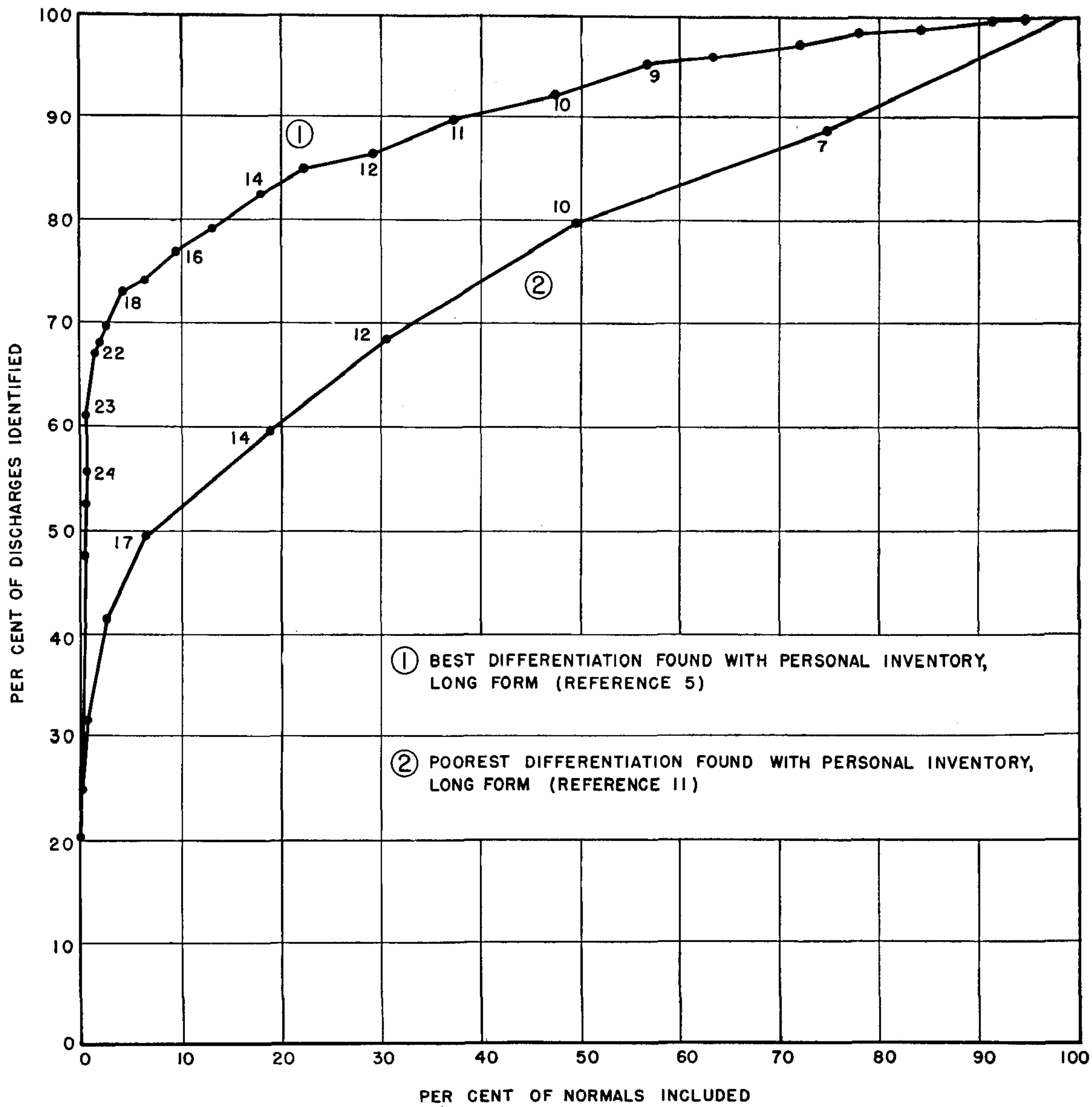


FIGURE 1. Value of Personal Inventory, Long Form, in differentiating psychiatric discharges from normals.

The two curves show the best and the worst differentiation found with the PI, Long Form. Each point, on each curve, represents a possible cutting score. Several such scores are indicated on each curve. The percentage of psychiatric discharges who equaled or exceeded each cutting score is shown on the vertical axis. The percentage of normals who equaled or exceeded each score is shown on the horizontal axis.

Curve 1 is based on 508 normals and 188 discharges. Curve 2 is based on 307 normals and 89 discharges.

represent different possible cutting scores. Each curve shows, along the horizontal axis, the percentage of a group of normal men who were included along with the percentage of psychiatric discharges indicated on the vertical axis. The two curves represent the best and the poorest differentiation found in any of the studies of the long form. Both are very much better than chance differentiation. In the best differentiation, 70 per cent of the psychiatric discharges were identified by a cutting score which included only 3 per cent of the normals. In the poorest differentiation, approximately 43 per cent of the discharges were identified by a cutting score which also included 3 per cent of the normals.

Expressing the results in terms of percentages gives an unduly favorable account of the usefulness of the inventory, for it neglects the fact that the satisfactory men far outnumber the unsatisfactory ones. At Newport Naval Training Station approximately 4 per cent of the men were given early psychiatric discharges. Thus about 40 were discharged of each 1,000 men examined. According to the best results shown in Table 1, approximately 28 of these 40 men could be identified by a cutting score which also picked up 29 of the 960 acceptable men in each 1,000. With less favorable discrimination the same cutting score might include considerably more normal than psychiatric cases. However, administration of the PI selects a small group (in the above case 57 men) which includes 70 per cent of all who would be discharged if the entire 1,000 were examined. Thus even though normal men were included, the amount of time necessary for psychiatric examinations could be greatly reduced without overlooking a very large fraction of those men whom the psychiatrist would discharge if time permitted the examination of all recruits. If time of the psychiatrist is at a premium, as it usually was during the war, the PI can help materially in the selection process.

4.2.2 Enlisted Men's Personal Inventory, Short Form[7]

After considerable experience with the long form of the PI a short form of 20 items was constructed. This form includes the ten most discriminative items from the long form and ten other items which showed fairly high discrimination between normal and psychiatric groups. These were included to give continuity in the subject matter of the short form. The 20 items are those marked with an (*S*) in the final column of Table 1. The short form is given in its entirety in Table 6.

TABLE 6. Personal Inventory, Short Form.

In each question, mark (X) the answer which fits you best. Even if neither answer fits you very well, mark the one that fits you better than the other.

1	I graduated from high school Yes ( )	1	No ( )
2	I was a sickly child ( )	2	I was an active child ( )
3	I have felt bad more from head colds ( )	3	I have felt bad more from dizziness ( )
4	I seek excitement ( )	4	I avoid excitement ( )
5	I like to have people do things my way ( )	5	I like to have people figure things out for me ( )
6	I am more nervous ( )	6	I am more easy going ( )
7	Somehow I never could find enough to do in my free time ( )	7	My free time always seemed to be filled ( )
8	I wish I wouldn't feel so tired ( )	8	I wish I could have a more responsible job ( )
9	I wish I could have more excitement ( )	9	I wish I weren't bothered by bad dreams ( )
10	I wish I didn't have so many aches and pains ( )	10	I wish I wouldn't keep changing my mind ( )
11	I wish I weren't so nervous ( )	11	I wish I wouldn't keep putting things off ( )
12	I wish I could get myself to take more chances ( )	12	I wish worrying wouldn't make me sick to my stomach ( )
13	I have more headaches than the average person Yes ( )	13	No ( )
14	The hours at night seem long Yes ( )	14	No ( )
15	I like most any kind of food ( )	15	I have a poor appetite ( )
16	After exertion I feel hungry ( )	16	After exertion I feel dizzy ( )
17	When excited I feel weak ( )	17	When excited I feel stronger ( )
18	I think I might like to watch a surgical operation sometime ( )	18	The sight of blood upsets me ( )
19	My heart sometimes speeds up for no reason at all Yes ( )	19	No ( )
20	I have never gone to a doctor for headaches or dizzy spells ( )	20	I have occasionally gone to a doctor for headaches or dizzy spells ( )

SCORING KEY

All items of the short form are scored. Table 7 shows the distribution of the critical ratios on two different populations. The correlation between the critical ratios determined in the two studies was .56.

TABLE 7. Critical ratios of the items in the Personal Inventory, short form.

Critical ratios	Number (first study)	Number (second study)
2.5– 4.9		1
5.0– 7.4	3	2
7.5– 9.9	7	4
10.0–12.4	6	7
12.5–14.9	1	5
15.0–17.4	3	1
	20	20

TABLE 8. Reliability coefficients (odd-even, Spearman-Brown corrected) for the Personal Inventory, short form.

Group	r_t
458 Newport recruits	.66
536 Newport recruits	.68
426 Amphibious Forces men	.89*
257 Newport discharges	.91

* This reliability coefficient was based on a first half–second half rather than an odd-even correlation.

RELIABILITY

The short form proved to be as reliable as the long form. Table 8 shows the results from four groups of subjects.

VALIDITY

The short form was shown to differentiate between normal men and discharges as satisfactorily as the long form.[9, 12] Curves similar to those shown in Figure 1 for the long form are presented in Figure 2 for the short form.

The data pictured in the upper curve of Figure 2 were obtained in a nonpredictive situation. The 263 discharges had already been so classified and the 538 normals had been accepted for duty. The data pictured in the lower curve were obtained in a predictive situation. The PI was administered and the answer sheets filed. Not until after the men had been accepted or discharged were the papers scored and the scores compared with the disposition of the men.

Comparison of the two curves in Figure 2, or the data in Table 9, shows less sharp differentiation in the predictive than in the nonpredictive situation. For example, using a cutting score of 7, it was found that 60 per cent of the discharges could be predicted correctly in advance of the psychiatric interview at a cost of including 9.6 per cent of the normals. Using the same cutting score in a nonpredictive situation,

TABLE 9. Cumulative percentage falling at and above each score on the Personal Inventory, short form, predictive and nonpredictive situations.

PI score	Predictive use: 458 normals N	Cum. %	Predictive use: 30 discharges N	Cum. %	Nonpredictive use: 538 normals N	Cum. %	Nonpredictive use: 263 discharges N	Cum. %
20	0	....	0	....	0	....	0	....
19	1	0.22	0	....	0	....	5	1.90
18	0	0.22	1	3.33	0	....	14	7.22
17	0	0.22	0	3.33	1	0.19	15	12.93
16	0	0.22	0	3.33	0	0.19	17	19.39
15	1	0.44	0	3.33	0	0.19	23	28.14
14	1	0.66	2	10.00	2	0.56	18	34.98
13	0	0.66	1	13.33	0	0.56	20	42.59
12	0	0.66	0	13.33	1	0.74	10	46.39
11	4	1.53	5	30.00	4	1.49	15	52.09
10	6	2.84	3	40.00	7	2.79	24	61.22
9	4	3.71	3	50.00	2	3.16	12	65.78
8	9	5.68	1	53.33	7	4.46	8	68.82
7	18	9.61	2	60.00	11	6.51	8	71.86
6	22	14.41	0	60.00	37	13.38	11	76.05
5	42	23.58	3	70.00	42	21.19	14	81.37
4	55	35.59	1	73.33	63	32.90	7	84.03
3	84	53.93	4	86.67	92	50.00	9	87.45
2	94	75.11	1	90.00	124	73.05	18	94.30
1	93	94.76	2	96.67	106	92.75	11	98.48
0	24	100.00	1	100.00	39	100.00	4	100.00

[redacted]

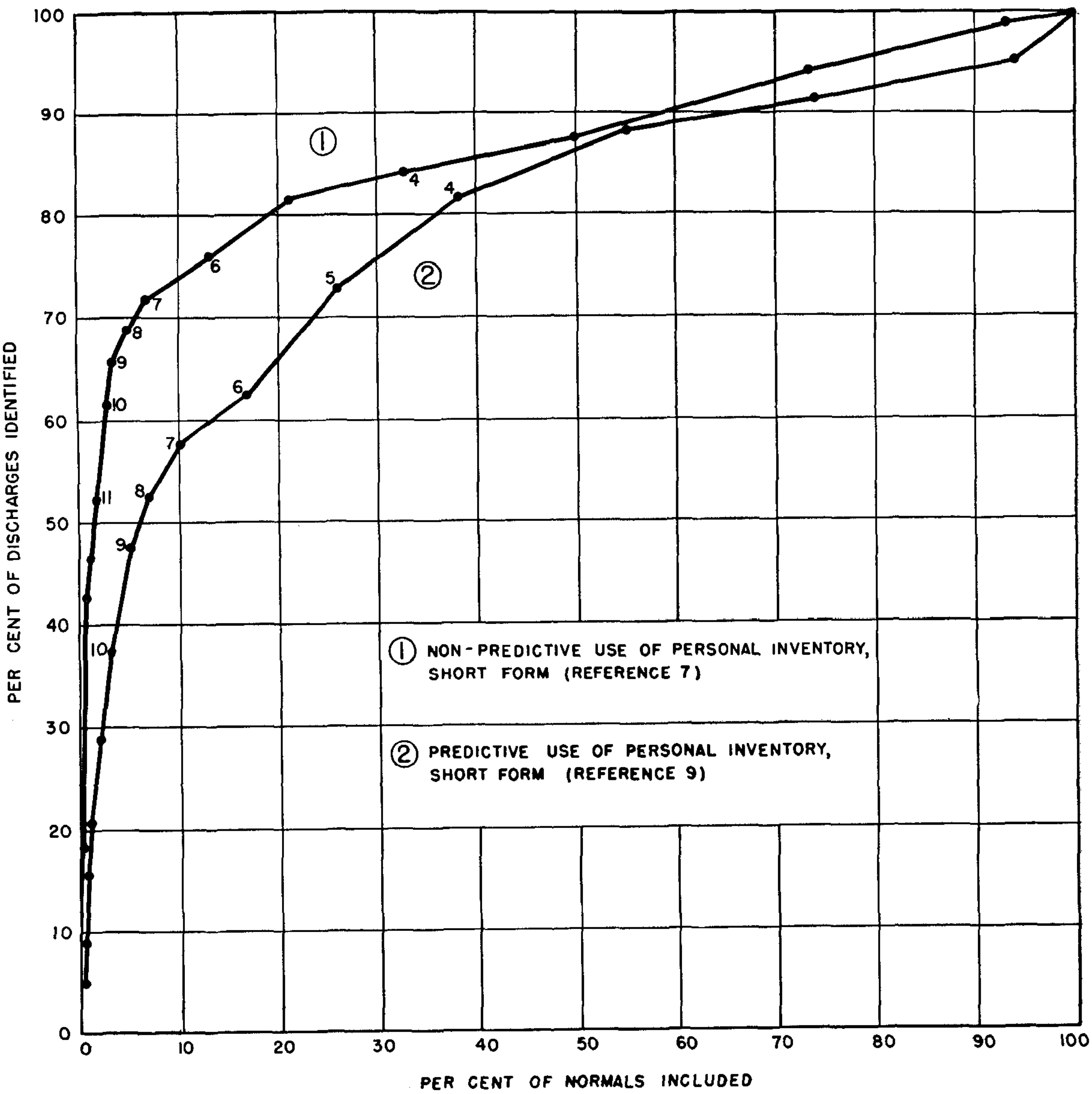


FIGURE 2. Comparison of discriminative value of Personal Inventory, Short Form, in predictive and nonpredictive use.

This figure is to be interpreted in the same way as Figure 1. The curve for predictive use is based on 815 normals and 59 discharges; that for nonpredictive use on 538 normals and 263 discharges. Better differentiation was found in the nonpredictive than in the predictive situation. Compare with Figure 1 for relative discrimination value of long and short forms of PI.

clearer differentiation between the two groups was achieved: 72 per cent of the discharges were identified at a cost of 6.4 per cent of the normals. But even the poorer differentiation shown in the predictive situation was as good as that shown in one of the nonpredictive studies using the long form (Figure 1). The usefulness of the short form in identifying an appreciable proportion of the psychiatrically undesirable men was unequivocally demonstrated.

4.2.3

Agreement between Long and Short Forms[11]

The relation between the long and short forms was shown by a correlation of .84 between scores made on the two forms by a group

[redacted]

of 426 men tested at the Amphibious Training Base, Solomons, Maryland. The correlation between the long and short forms was essentially equal to the reliability of each. Furthermore, the two forms were equally valid. Choice between them became, therefore, largely a matter of preference. Both the long and the short forms continued in use throughout the war, with the Coast Guard and Maritime Service using the long form and the Navy adopting the short form (see Section 4.4).

4.2.4 Comparison with GCT in Identifying Psychiatric Cases[11]

The General Classification Test (GCT) was already given to all Navy recruits. It was, therefore, important to determine whether the PI worked more efficiently in selecting men who would be given psychiatric discharges than did the GCT.

TABLE 10. Correlation between GCT and PI scores.

Group	r
1,071 Newport recruits	—.28 (old form GCT)
89 Newport psychiatric discharges	.01 (old form GCT)
400 Submarine School candidates	—.17 (Otis test)
576 Amphibious Forces enlisted men	—.35 (new form GCT)
426 Amphibious Forces enlisted men	—.25 (new form GCT)
412 Newport recruits	—.28 (old form GCT)

The correlation between PI scores and GCT scores varied from .01 to —.35. (The negative correlation was due to the fact that high scores on the GCT and low scores on the PI were considered "good.") The actual correlations for six groups are shown in Table 10.

The PI was found to be more efficient than the GCT at any cutoff score in discriminating between normal men and those given psychiatric discharges. Table 11, for example, demonstrates that a score of 72 on the GCT identified 60.7 per cent of the psychiatric discharges but included 42.4 per cent of the normals. A score of 14 on the PI identified 59.6 per cent of the discharges and included 18.9 per cent of the normals. Discrimination by the PI was obviously superior to that obtained by the GCT.

The figures in Table 11 were obtained with the long form of the PI. No direct comparisons of the short form with the GCT are available. However, since the short form was found to correlate .84 with the long form, and since the two forms were shown to be approximately equal in validity and reliability, it seems highly probable that the short form, like the long one, is considerably more discriminative of psychiatrically undesirable men than is the GCT.

4.2.5 Validation against Other Criteria

All validation discussed above was obtained against the criterion, psychiatric judgment. It would be highly desirable to be able to present validation against criteria of combat proficiency or to use as criteria other records

TABLE 11. Comparison of identification of discharges by GCT (old form) score and PI score. For normals, $N = 307$; for discharges, $N = 89$.

GCT cutoff score	Percentage discharges included	Percentage normals included	PI cutoff score	Percentage discharges included	Percentage normals included
34	10.1	0.7	37	10.1	0.0
43	20.2	1.3	29	20.2	0.0
51	30.3	6.2	24	31.5	0.3
59	38.2	14.3	20	41.6	2.3
67	50.6	28.7	17	49.4	6.2
72	60.7	42.4	14	59.6	18.9
75	70.8	49.8	12	68.5	30.3
81	79.8	66.8	10	79.8	49.8
87	91.0	86.0	7	88.8	74.6
97	100.0	99.4	2	100.0	99.7

[redacted]

indicating how satisfactorily each man withstood the pressures of military training and activity. Such criteria were extremely difficult to secure in any reliable form. An effort was, however, made to validate the PI against criteria other than the psychiatrist's initial judgment.

Comparison with Service Records[10]

One study compared PI scores with Service records one year after testing. The data are summarized in Table 12. The object of this study was to relate PI scores to the rating held one year after testing, to the conduct record during the first year of enlistment, and to the active or discharge status one year after testing. The records of 1,466 men were examined. All had been accepted by the psychiatrists at the time of original testing and boot training. Complete records were available on 1,007 of these men. Separate analyses of the total group of 1,466 and of the 1,007 men for whom complete records were available were in all respects practically identical. The principal results of the analyses of the total group were as follows:

1. The PI identified a significant proportion of the 52 men who were later discharged. Twenty-one per cent of these men had received scores of 18 or above on the PI, as compared with but 4 per cent of those not discharged. The mean score for the discharges was 4.3 points higher than for the active men (critical ratio 3.9). Of the seven men who scored 26 or over on the PI, all were discharged during the first year of service.

2. The PI showed some tendency to differentiate between good and bad conduct cases. The mean score for the men with good conduct records was 1.7 points lower than for the men with some conduct offenses (critical ratio 4.1), and 2.8 points lower than for those with more serious offenses (critical ratio 4.6).

3. The PI showed some tendency to differentiate rated from nonrated men. The mean score for the rated men was 1.7 points lower than for the nonrated men (critical ratio 6.1).

4. The GCT tended to differentiate discharges from those still on active duty, but to a decidedly lesser extent than did the PI. It tended to differentiate conduct cases to a slight degree. It differentiated rated men from unrated ones somewhat more sharply than did the PI. This may have been due in part to the availability of the GCT scores in the assignment of ratings.

While this evidence shows some relations between PI scores and records one year later of conduct, rating, and discharge status, PI scores would have little value in selecting men for advancement. The restricted range of scores in this population should, however, be remem-

TABLE 12. Relation between Navy Service records and scores on the PI.

Group	N	Mean	σ	Sigma units of mean from general mean	Score of 18 and above N	Score of 18 and above %
Entire group	1,466	9.3	4.6		63	4.3
Nonrated men	856	10.0	4.8	+0.15	51	6.0
Rated men	609	8.3	4.2	−0.22	12	2.0
Nonoffenders	1,328	9.1	4.6	−0.04	52	3.9
Offenders	138	10.8	4.8	+0.33	11	8.0
Worst offenders	64	11.9	4.7	+0.57	7	10.9
Active	1,414	9.2	4.4	−0.02	52	3.7
Discharges	52	13.5	8.1	+0.91	11	21.2

Critical ratios of differences between means	
Nonrated vs rated	6.1
Nonoffenders vs offenders	4.1
Nonoffenders vs worst offenders	4.6
Active vs discharges	3.9

[redacted]

bered in considering these data. All men included in the study had been approved by the examining psychiatrists. What results would have been obtained on an unselected population is unknown.

Use of the PI in Selecting Paratroopers[15]

A second investigation using a criterion other than psychiatric judgment studied the usefulness of the PI in predicting success or failure in the parachute school at Fort Benning, Georgia. Of 1,079 men tested there with the PI, 778 successfully completed parachute training. The remaining 301 men fell into 13 categories of failure. The four largest failure categories were: 89 men classified as having insufficient desire to continue training; 60 men who refused to jump from mock-up towers; 34 men who were permanently disqualified physically; and 67 men who were injured in training and temporarily disqualified for varying lengths of time. Some of these would be expected to complete training later and others would later be classified as permanently disqualified physically.

The biserial correlation between PI score and completion or failure to complete parachute training was —.39. The correlation for the Army General Classification Test was .26. (The population was a normal Army group in terms of Army General Classification Test scores.) These two correlations can be combined into a multiple correlation of .41 with success in parachute training. On the basis of these findings, it appeared that the PI could contribute to the improved selection of men for parachute training, and that the PI would be as effective for this purpose alone as in combination with the Army General Classification Test.

Escape Tank Training[4]

A third attempt to validate the PI against a performance criterion used success in escape tank training at the Submarine School, New London, Connecticut, as a criterion. The submarine escape training tank is a circular metal tank 100 feet in height filled with fresh water at approximately 90 F. Men reporting for training in the use of the submarine escape appliance, the "lung," are first given a brief physical examination. They are then placed in a decompression chamber where they are instructed in the principles on which the lung operates. After the pressure is reduced to normal, the men are given lungs and further instructions in their use. The training itself consists of going down a ladder into the water and using the lung. Later each man is taken down to a depth of 12 feet in a diving bell. He is required to ascend to the surface on an escape line twice. Finally he is required to make two satisfactory escapes from a lock at an 18-foot depth. Many men also escape from the 50- and 100-foot locks, but this is not required. About 1½ per cent of the men appeared to be seriously disturbed emotionally. This disturbance was shown by letting go the line and rising rapidly to the surface; spitting out the mouthpiece of the lung; coming up too fast; refusing to leave an escape chamber; excessive complaining or other signs of emotional disturbance.

The PI was found to have a tetrachoric correlation of .14 with tank failure. In combination with other tests, a multiple correlation of .43 appeared, but the PI alone could not be used satisfactorily to predict success or failure in tank escape training.

Selecting Submarine Personnel[18]

The PI and a number of other tests were administered to all men processed at the New London Submarine Base in 1943 and the early part of 1944. These test scores were later compared with trait ratings of the men made by submarine officers.

Scores on the PI, and on the other selection tests also, were found to be unrelated to the quality of performance aboard submarines as it was rated by officers. The absence of correlation can be attributed to the unreliability of the criterion and to the fact that the sample of men studied was already selected with respect to mental, physical, and emotional characteristics.

Under these same conditions, the psychiatrist's evaluations of men judged to be emotionally unstable were no more useful in predicting the criterion than were scores on the PI.

4.2.6 Officers' Personal Inventory

CONSTRUCTION

One form of the PI was constructed for use with officers. This form consists of 164 items. Forty-nine items were taken from the Long Form of the enlisted men's PI with some changes in the wording of some items; the other 115 were new. In general format and in scoring method it was similar to the Long Form of the enlisted men's PI. The officers' PI is reproduced in the appendix.

SCORING KEY[13]

Arrangements were made for a validational study of the officers' PI in the Amphibious Training Command, U. S. Atlantic Fleet. Two conditions made validation impossible. In eight months' time only 83 officers were diagnosed as emotionally disqualified by the Boards of Medical Examiners at the Amphibious Training Bases. Further, these Boards insisted on having access to the tentative PI scores at the time of their own examinations. The PI scores and the medical examiners' ratings were therefore not independently arrived at. Consequently, the ratings could not be used as a basis for validating the PI scores.

The scores supplied to the Boards of Medical Examiners were based on a tentative 34-item scoring key. These 34 items are identical with ones contained in the scoring key of the long form of the PI used for enlisted men. Since the medical examiners did not have access to the individual answer sheets, they did not know how any man had answered the individual items. It was, therefore, possible to study the ability of each of the 164 items to discriminate between the 83 emotionally disqualified officers and a normal group of qualified officers. Thirteen hundred officers were included in the normal group.

On the basis of item analysis comparisons between these two groups, a scoring key of 50 items was constructed. Each of the 50 items showed a critical ratio of 3.96 or greater in differentiating between the emotionally disqualified and the qualified officers of this study. A cutoff score of 30 on this key identified 53 per cent of the emotionally disqualified officers while including 2.6 per cent of the qualified officers. The odd-even reliability of this key is represented by a correlation coefficient of .88 corrected by the Spearman-Brown formula. Further study involving the application of this key to new groups of emotionally qualified and disqualified officers will be needed to establish its validity. In such further study the diagnosing medical officers must be without knowledge of the inventory scores at the time of diagnosis if validity of the PI for officers is to be established.

PREDICTING SUCCESS IN MARINE CORPS OFFICER CANDIDATE SCHOOL

A study[20] of the officers' PI was conducted by the Medical Field Research Laboratory at Camp LeJeune, North Carolina, under better controlled conditions than were possible in the Amphibious Training Bases. At Camp LeJeune the officers' PI was administered to 1,039 Marine Corps officer candidates. Each man was tested within one week after arrival at Camp LeJeune for a period of military training and observation for screening purposes. All men had V-12 college training and all had just completed 12 weeks of recruit training at Parris Island. A follow-up study was made on the subjects after they had either completed or failed to complete officer candidate school four or five months later. Test results were seen by no one not connected with the research staff.

A scoring key was constructed using the items which best differentiated between those who completed and those who failed to complete officer candidate school. Using this scoring key on the same group of men on whom it was constructed showed a biserial correlation of .48 between PI scores and success or failure in officer candidate school. When this scoring key was applied to new groups of 757 candidates[21] and 671 candidates,[22] correlations of .18 and .28 were found. For the two groups combined the correlation was .25. On these same groups the original NDRC-Navy scoring key correlated with the pass-fail criterion to the extent of .22 and .09. For the combined groups, the correlation was .16. These correlations indicate that the PI would not serve effectively as a single basis for selecting Marine Corps

[redacted]

officer candidates. It may prove useful as part of a selection battery, however. Tentative data on this possibility were secured by the Medical Field Research Laboratory, Camp LeJeune, North Carolina, which found a multiple correlation of .32 between the PI and a specially prepared confidential questionnaire and a pass-fail criterion in officer candidate school.[23]

4.3 OTHER METHODS OF SELECTING EMOTIONALLY UNSTABLE MEN

4.3.1 Tests Similar to the Personal Inventory

NEW LONDON-NDRC QUESTIONNAIRE

Reference 3 gives an analysis of the data obtained from testing 417 normal and 71 psychiatric ward cases at the Newport Naval Training Station by means of the New London-NDRC Questionnaire developed under an NDRC Division 7 project. This questionnaire did not appear to be as effective as the PI in distinguishing between acceptable and unacceptable men. Information concerning this questionnaire and its trial use at the U. S. Submarine Base, New London, Connecticut, is contained in reference 2.

THE CORNELL SELECTEE INDEX

The Cornell Selectee Index was developed under the Committee on Medical Research of OSRD. It used a different type of question (e.g., Did you ever regularly drink more than three quarts of whiskey a week?). A 32-item edition of the Cornell Selectee Index and the short form of the PI were tried out under identical conditions on 1,000 recruits at Newport Naval Training Station. Both tests were taken before the recruits were given psychiatric examinations, but the scores were not available until after psychiatric disposition of the men was complete. In this purely predictive situation, the PI identified 61 per cent of the discharged men at a cost of falsely identifying 14 per cent of the normals. The Cornell Selectee Index picked up 67 per cent of the discharges and 19 per cent of the normals. Using the two tests together resulted in the correct identification by one or both of them of 82 per cent of the discharges and the false identification of 24 per cent of the normals. [Letter from Commanding Officer, U. S. Naval Training School, Newport, Rhode Island, to Chief, Bureau of Medicine and Surgery, March 3, 1944, Serial No. 14939, P2-5 (1A)].

The Enlisted Personal Inventory, Form 2, NavPers 16845, adopted for general Navy use, contains items from the Cornell Selectee Index in addition to the 20 items of the short form of the PI.

THE PERSONAL INVENTORY, ARMY AIR FORCES REVISION

As part of the aviation psychology research program in the Army Air Forces, the Personal Inventory was revised for use in classification and redistribution of AAF officers and gunners. The history of the Aviation Psychology Program[24] contains an account of the development of several modified forms of the PI and of the satisfactory use of these modifications.

4.3.2 Use of Battle Noise Equipment in Screening Psychiatrically Undesirable Personnel[14, 17]

Section 17.3 of NDRC developed, at Navy request, an extremely high-power, high-fidelity sound reproducing system. Upon completion, this equipment, the Battle Noise Equipment, was assigned to the U. S. Naval Training Station, Newport, Rhode Island. The medical officers there were directed to study its potential usefulness in screening out those men who would be so disturbed emotionally by the noise of battle that they would be ineffective in combat. Assistance of the Applied Psychology Panel in conducting these studies was requested.

Sound records of battle action were played at intensities of 110 db and higher to groups of men whose reactions to the noise were studied. During the playing of any sound sequence the men were observed by a psychiatrist who watched for any obvious signs of disturbance. In some cases the signs were very obvious. A few men broke ranks and ran away.

A few were taken directly to the psychiatric ward. Many more showed pallor, sweating, or other emotional signs. At the end of the noise exposure each man was asked to indicate, by holding up his hand or by writing on a card, which of a list of symptoms described by the experimenter he had experienced.

Observations of this kind were made on normal recruit groups and on men from the psychiatric ward. The sound was sometimes presented alone and sometimes as an accompaniment to motion pictures of battle scenes. The motion pictures were shown to some men without noise. Finally, some men were not exposed to the pictures or the noise but were merely asked to check those symptoms they would feel if they heard the noises of battle.

In each case the men were classified in accordance with the number of symptoms exhibited or reported. The resulting classification was then compared with the diagnosis arrived at in the regular psychiatric interview.

It was found that the men's reactions to the battle noise, either alone or with motion pictures of battle scenes, agreed fairly well with the results of the psychiatric interview. But it was also found that the men's responses to the questions alone, without ever hearing the battle noises at all, served just as effectively in picking out those whom the psychiatrist considered unfit for duty.

Consequently no change in the psychiatric procedure was recommended and the battle noise equipment was not used for screening purposes. The Applied Psychology Panel recommended to the Navy that the equipment be used to give increased realism to some types of attack training. That recommendation was adopted and the equipment put to use in the Marine infiltration range at Camp Pendleton, California.[16]

4.4 SERVICE USE OF THE PERSONAL INVENTORY

4.4.1 Possible Uses

The Personal Inventory might be used by the military Services in either of two ways: as a method of selecting men for special types of duty, or as a psychiatric screen for picking out those who should be given more detailed examination. In view of the criteria which were available for validating the PI, there was little justification for using it as the basis for selecting men for special duty without further study. There was much justification for using it as a psychiatric screen.

In many installations the medical staff was not able to give thorough psychiatric interviews to every man passing through the installation. In such situations the PI provided a method of picking out a small group which contained most of the men who would have been discharged if time had permitted the careful examination of all men. Results obtained by using the short form of the PI in this fashion are presented in Table 13. This table is based on experience with a wartime recruit population in which about 5 per cent of the men were given psychiatric discharges. At other times, and with other groups, the absolute figures would differ throughout the table. But the principle of saving time by selecting for interview those men most likely to be rejected is applicable generally.

Table 13 shows the saving in interview time which can be achieved by using the PI as a preliminary screen. For example, at a cutting score of 7, 156 men out of each 1,000 examined would be identified for interview. Of these 156 men, 32 would be discharged. These 32 constitute about 60 per cent of the total number of men who would be discharged had the whole 1,000 been examined. Thus 60 per cent of the potential discharges can be identified by a process which requires interviewing only 16 per cent of the entire population. If the psychiatric examining facilities allow careful examination of a larger fraction of the total flow of men, a lower cutting score can be adopted. For example, a score of 5 would require examining 308 of the 1,000 men. That group of 308 would contain 70 per cent of all who would be discharged if the whole 1,000 were examined. Using the information contained in Table 13, it is possible to adjust the cutting score in terms of the total number of men passing through a station and the time available for psychiatric examinations.

[redacted]

4.4.2 Psychiatric Endorsement

Use of the Personal Inventory was endorsed by the Subcommittee on Psychiatry of the National Research Council's Division of Medical Sciences at a meeting on June 29, 1944, in a resolution which contained the following recommendations:

"NOW THEREFORE BE IT RECOMMENDED that these methods be put to use as widely as possible in the military services, and that further studies along these lines be continued and extended with a view to the progressive improvement and refinement of these tests in the light of validating data from actual performance under conditions of military service and combat;

BE IT FURTHER RECOMMENDED that tests of a similar nature be developed for use in the selection of personnel for all types of special assignment."

Table 13. Efficiency of the Personal Inventory in reducing the number of men to be interviewed.

Cutting score	Number identified for interview per 1,000	Number interviewees to be discharged	Number not interviewed, who would be discharged had they been interviewed
20	0	0	53
19	2	0	53
18	4	2	51
17	4	2	51
16	4	2	51
15	5	2	51
14	12	5	48
13	16	7	46
12	19	7	46
11	39	16	37
10	56	21	32
9	79	26	27
8	110	28	25
7	156	32	21
6	212	32	21
5	308	37	16
4	422	39	14
3	594	46	7
2	786	47	6
1	956	51	2
0	1,000	53	0

4.4.3 Service Adoptions of the PI

The PI was used experimentally in a number of places by the Navy, Coast Guard, U. S. Maritime Service, Army Air Forces, Army Ground Forces, and the Marine Corps. One or more forms were adopted for routine screening purposes by the U. S. Navy, Coast Guard, and Maritime Service. A special form was developed by the Army Air Forces[24] (see Section 4.3.1).

Format A (the long form with a scoring system giving separate scores on each of 20 clusters of items) was adopted as part of the routine selection procedure at the U. S. Submarine Base, New London, Connecticut. On May 22, 1944, a directive from the Bureau of Naval Personnel specified a score of 14 or lower on Format B for admission to Submarine School.

Format B (the long form with a simpler scoring procedure giving only one score) was adopted as part of the routine selection procedure at the U. S. Navy Intercept Officer Training School at St. Simons Island, Georgia; at the U. S. Maritime Service Training Stations located at Sheepshead Bay, New York; Catalina Island, California; and St. Petersburg, Florida; and in the U. S. Coast Guard classification centers.

On January 24, 1945, the Director of Enlisted Personnel wrote to the commanders and commanding officers of some 50 receiving stations,

naval training centers, and training schools informing them that the short form of the PI (Enlisted Personal Inventory, Form 2, NavPers 16845) "has been adopted by this Bureau to aid in identifying enlisted personnel who lack the emotional stability required by submarine and amphibious duty." "Men making scores of 8 or higher . . . are ineligible for transfer to Submarine School, New London." Instructions were given to make careful psychiatric examinations of all men in the Amphibious Forces making scores of 10 or above.

The Officers' Personal Inventory was not adopted for routine use in any Service.

4.5 RECOMMENDATIONS FOR FURTHER WORK

The PI worked more successfully in military selection than it will in selecting civilians for peacetime jobs. On all personality tests of this type it is possible for a man to make a higher or a lower score than he deserves, for the normal, nonsymptomatic answers to the questions are sometimes fairly obvious. The most probable civilian use of the PI would be in selecting men for positions for which the men themselves apply. Under these conditions there will be considerable temptation to give "normal" answers to the questions and thus to make a better score than is deserved.

In wartime the motivation is to a large extent reversed. Some men answer the questions in such a way as to make a poorer or less normal score than they otherwise would, hoping thereby to avoid military duty and particularly the more hazardous types of duty. The PI, consequently, lumps together those who are really emotionally unstable and those who are willing to admit more psychiatric symptoms than they normally exhibit. Since both of these groups should be examined in detail by the psychiatrists, this grouping is not a serious fault in military use of the PI.

The distinction between wartime and peacetime conditions of use of this test is pointed out here to warn against any expectation that the success of the PI in screening out men unsuitable for military duty can be duplicated in peacetime industrial situations. How well it will work in peacetime military situations remains to be seen.

Efforts to construct personality tests which are completely free from the possibility of faking have met with little success. Nevertheless these efforts should be continued. Tests of emotional stability and personality traits could be used with much greater confidence if the men being tested were unable to secure better scores than they deserve by giving false answers to the test items.

The PI was originally intended as a means of predicting a man's ability to perform satisfactorily in hazardous duty or in situations requiring great emotional stability. In practice it became a device for predicting what a psychiatrist would say about a man on the basis of a short interview. A thoroughgoing validation of the psychiatrist's own judgment is needed. Future work on instruments similar to the PI should be accompanied by parallel studies of the psychiatrists' judgment. Both of these methods of predicting a man's emotional stability should be validated against an independent criterion, preferably one based on combat performance. The difficulty of securing such criteria is great, but that difficulty must be overcome if instruments similar to the PI are to develop beyond the stage of being predictors of the unvalidated judgment of the psychiatrist.

[redacted]

Chapter 5

DETERMINING VOCATIONAL INTERESTS

By *Dael Wolfe*[a]

Summary

An inventory of interests was constructed for possible use in the classification of military personnel. The inventory was intended to secure information on a wide variety of interests, e.g., vocational, social, recreational, scientific, and religious. Information of this type could provide a valuable supplement to that obtained by interviews, vocational history, or standarized tests of ability, in classifying men for military duty.

The interest inventory, a scoring key, and detailed statistical information were turned over to the War Department Adjutant General's Office. The inventory has some advantages over other existing interest questionnaires. It has not yet been validated for military personnel.

5.1 INTRODUCTION

Information regarding a man's interests combined with information about his ability ought to provide a better basis for job assignment than would either type of information alone. Previous attempts to secure information about interests in standardized and comparable form were not considered satisfactory for military use. The Adjutant General's Office therefore requested the Applied Psychology Panel to construct an interest inventory for use in the classification of military personnel. The work was assigned, as Project SOS-7, to a contract with Harvard University.

5.2 THE ACTIVITY-PREFERENCE TEST

5.2.1 Construction of Items

Several paper-and-pencil tests of interest were already in existence at the time this project work began. The two most widely used were those developed by E. K. Strong and G. F. Kuder. Both attempted to determine a man's interests by asking a number of questions about the activities which he liked or disliked. Typical of the *Strong Vocational Interest Blank for Men* was a list of related items (such as hobbies or school subjects). The subject marked each to show whether he liked, disliked, or was indifferent to that activity (hobby, school subject, etc.). The *Kuder Preference Record* consisted of paired items, or groups of three items. The subject indicated the most and least preferred activity within each group.

The Activity-Preference Test differed from both earlier tests in several respects, although it continued the tradition of attempting to determine general types and fields of interests by asking a series of questions on quite specific likes and dislikes. It differed in attempting to cover a wider range of activities; in attempting to make the items compared more directly competitive with each other in terms of time, cost, and opportunity than was true of the earlier tests; and in including questions on biographical background and attitudes toward oneself.

The Activity-Preference Test contained a total of 220 *exercises*. Each exercise, except for items of factual information included in Section I, consisted of four alternative statements, each of which described an activity or situation. Taking the test required selecting the most liked and the least liked one of the four activities described, or selecting the most likely and the least likely behavior or response under the conditions stated. The item form was, therefore, similar to that used in the *Kuder Preference Record*. Two examples are:

Upon a free afternoon, which of the following would you like MOST and which would you like LEAST?

Exercise 74:

A. to go to a vaudeville show
B. to help organize a Boy Scout Troop
C. to take part in a play
D. to take the family out for a picnic

Think of some situation in the past few years in

[a] This chapter is based entirely on the final report of Project SOS-7.[1]

[redacted]

which you have succeeded markedly. Which of these do you think was MOST and which do you think was LEAST responsible for the success?

Exercise 22:

A. I always worked hard to earn quick money.
B. I had an excellent memory for people's faces.
C. I was not timid in making constructive suggestions.
D. I put more energy than most people into getting a job done.

The 220 exercises were divided into seven sections.

Section I (10 exercises) covered biographical information concerning the person when he was 15½ years old.

Sections II and V (15 exercises each) covered preferences and interests as they existed at age 15½.

Sections III and VI (75 exercises each) covered preferences at the time of taking the inventory.

Sections IV and VII (15 items each) required the subject to look ahead to the time when he would be 48 years old and to indicate the activities which he expected to prefer at that age.

Two forms of the inventory contained the same exercises, but in different orders.

The *like most–like least* form required two responses to every exercise and thereby avoided the systematic error consequent to a rosy or blue outlook which in other interest tests has resulted in excessive marking of things as being "liked" or as being "not liked." There may be a "general liking of things" function, and in some connections it may be important. The present instrument had evidence upon this only in the first, or biographical, section. Even though one person "likes most everything" more than a second, still the lesser level of intensity of feeling of the second person is the intensity level that dictates his conduct. His contrasting interests at the low level may well be just as important in determining his vocational preference as the more intense level of the first person. A choice must be made, no matter at what level.

The four options within each exercise were very carefully devised so that for the average person there is no moral issue, or right or wrong, or good or bad choice, involved. Any expression of interest can be falsified, but if the subject sees no reason for dissimulation, and if there is no earmarked "right" or "wrong" answer, an honest expression of conviction can be expected. It is true that an intelligent person can produce an answer sheet which is biased in a direction which has been chosen by the person if he is so inclined. Such bias is not as great as might be expected because of the subtlety of the scoring schemes. The internal evidence suggested that a useful falsification—useful in the sense desired by the subject—of the activity-preference record is measurably less with this instrument than with any of the other commonly available interest tests. This difficulty of falsification was accomplished by the form employed and by the care directed to this issue in the construction of exercises.

As a general interest measure serving all subjects, the exercises were devised to sample as nearly as possible all fields of life in which preferences can be shown. A careful scrutiny was made of the waking activities of the entire day and of the seasons, spring, summer, fall, and winter; of the social contacts of life from the narrowest social group, the home, to the widest; of the earliest age (chosen as age 15½) involving the maturity and initial outlooks of adulthood, and of the last age presumably thought of as still in full prime (chosen as age 48). The results of these surveys were reflected in the wide variety of exercises included in the inventory.

It was also set down as a requirement that the four options constituting an exercise should be genuinely competitive. A subject was not asked to express a preference between a trip to Mexico, requiring much time and money, and seeing a football game, or to express a preference between swimming and skating, or between other activities which in the nature of things do not simultaneously claim attention.

5.2.2 Types of Interests Covered

The 220 exercises were planned to secure information on 40 different types of interests (rubrics). These were:

[redacted]

Rubric No.	Description
1	Things and mechanisms (not a repetitive or routine use).
2	Spatial thinking (visual-mindedness).
3	Calculation, computation, number facility (not symbolic thinking).
4	Orderliness, system, organization, regimentation.
5	Routine, undisturbing activities vs desire for change, variety, and suspense.
6	Monetary motivation and personally gainful activity.
7	Masculinity-femininity.
8	Family and neighborhood.
9	Adventure, daring, risk vs self-security, timidity.
10	Power (the leadership involving or based upon control) vs submission and satisfaction in position of menial servitude.
11	Gross physical activity in work and/or play.
12	Fine dextrous activity and craftsmanship.
13	Gregarious—bold and forward vs shy.
14	Handling people for their own presumed good.
15	Independent, self-sufficient vs gregarious.
16	Memory activity.
17	Pioneering and initiative (leadership based thereon) vs traditionalism.
18	Instructing, advising, supervising.
19	Salesmanship (not over-the-counter).
20	Verbal expression—oral and written.
21	Vocational level from professional to labor.
22	Verbal reception: reading and hearing.
23	Competitiveness.
24	Humor.
25	Music: participation and appreciation.
26	Arts: graphic, plastic, artistic: participation and appreciation.
27	Abstract vs concrete thinking.
28	Experimental verification vs fiat and a priori conviction.
29	Principal and suspended judgment vs impulse, instinct, sensuousness.
30	Practical reality vs fantasy.
31	Acting.
32	Religion and idealism.
33	Self-confidence vs tendency to worry.
34	Emotionality vs stolidity.
35	Perseverance and industry vs laziness and shillyshallying.
36	Outdoor vs indoor work, play, and amusement.
37	Plants and animals.
38	Integrated behavior vs neuroticism and compensatory adjustments.
39	Aesthetic and synaesthetic responsiveness.
40	Vehement, explosive, violent vs restrained, calm, quiet.

An a priori set of scoring weights was made up which permitted the tabulation of scores on each of these 40 rubrics. Each rubric score was obtained by totaling the weights assigned to each answer of each exercise considered appropriate for that rubric.

5.2.3 First Administration

The Adjutant General's Office arranged to administer the Activity-Preference Test to groups of enlisted men at three Army training centers. A total of 3,734 men was tested. Papers from a subgroup of 2,201 men all born in the years 1918 to 1925 were used for subsequent statistical analysis.

5.2.4 Item Study

Reliabilities of the rubric scores were computed using a priori weights for the data from Sections III and VI (total of 150 exercises at present age). The reliabilities of the separate scores ranged from .00 to .79. Eliminating the least reliable rubrics reduced their number to 34. The range of uncorrected reliabilities for the retained rubrics was .26 to .79.

Refined sets of scoring weights were obtained for 15 rubrics by analyzing the responses to each exercise scored for each of these 15 rubrics. The reliabilities of the scores based on refined weights for the exercises in Sections III and VI ranged from .50 to .84.

Estimated rubric reliabilities for the entire inventory ranged from .55 to .89, with a median of .75.

5.2.5 Factor Analysis

The variances for the 34 retained rubrics were adjusted in terms of the judged importance of each. The covariances were modified accordingly. The matrix of these adjusted variances and covariances was analyzed by the method of principal components. The matrix was analyzed into 34 components, the first five of which accounted for 72.31 per cent of the total variance.

The full meaning of these components depends upon their relations to success in a variety of occupations. Any formal naming of them is

[redacted]

difficult and likely to be misleading until a large amount of knowledge of their vocational relations and usefulness accumulates. The designations listed below were intended only for purposes of identification. Following each, the component variance is given, and then an indication of the predominant rubrics or personal characteristics important in each component.

MIMSEC, the first component (36.41)
M Masculine
I Isolationistic (prefers being alone)
M Mechanical

S Social
E Effeminate
C Conversational

PEPGAP, the second component (14.60)
P Persevering and pioneering
E Economic
P Practical and interest in power

G Gross (rough or boisterous, as opposed to refined)
A Adventurous and daring
P Physical activity (interest in)

PAMRIM, the third component (10.37)
P Power and gross physical activity (interest in)
A Aggressive (salesmanship and competition)
M Money (interest in)

R Religious
I Industrious
M Music, mechanisms, and craftsmanship (interest in)

RODPOD, the fourth component (5.71)
R Routine and religion (interest in)
O Orderly
D Domestic

P Pioneering (interest in)
O Outdoors (interest in)
D Daring

NEVCOM, the fifth component (5.23)
N Nature-loving
E V } EVangelistic—religion with salesmanship

C O M } Interest in general COMpetence, that is, positive weightings of power, vocational level, mechanisms, spatial, orderliness, verbal reception and expression, and music.

The names given to the components were each of two syllables which are equally well pronounced together if reversed, the first related to the rubrics having positive weightings in the component and the second to those having negative weightings. A person scoring high on the first component may be called a *mimsec,* and one scoring low a *secmim* person. The letters *mim* are related to rubrics with positive weightings, and the letters *sec* to rubrics with negative weightings. To those initiated into the nature and uses of this preference instrument, these letters may provide reasonable but very tentative pegs around which to build a more complete picture, and to the uninitiated (in general the person taking the test for the first time) these words will be harmless in that none of them suggest a type of "expected" or "right" response.

5.2.6 Preparation of Scoring Key

A final scoring key for each of the five principal components was constructed. Each item in Sections III and VI was assigned a weight within the range of —4 to +4, and each item in Sections I, II, IV, V, and VII a weight within the range of —6 to +6. For any one of the five components, most of the item weights were zero.

The reliabilities of the five component scores are shown in Table 1. These reliabilities were determined by correlating scores on one half of the indicated section or sections with scores on the other half of the indicated section or sections and then stepping up by the Spearman-Brown formula.

TABLE 1. Reliabilities of component scores.* ($N = 200$).

Components	Sections of test I	II & V	III & VI	IV & VII	Total test
I MIMSEC	.77	.87	.86	.79	.94
II PEPGAP	.54	.67	.83	.56	.86
III PAMRIM	.80	.76	.87	.51	.91
IV RODPOD	.83	.83	.78	.53	.89
V NEVCOM	.76	.61	.61	.55	.77

* It was estimated that with refined weightings for the items in Sections III and VI, the reliabilities of the several components for the entire instrument would be .96 for Component I; .89 for Component II; .93 for Component III; .89 for Component IV; and .81 for Component V.

The reliabilities are high enough to justify separating the inventory, for some purposes, into two equivalent tests, each half as long as the original. The estimated reliabilities of a half-length test for each component score would be:

MIMSEC = .89
PEPGAP = .75
PAMRIM = .84
RODPOD = .80
NEVCOM = .63

[redacted]

5.2.7 Final Disposition

The test, a table of scoring weights for the five components, and detailed statistical information concerning the test and the factor analysis were submitted, by request, to the Office of the Adjutant General at the end of 1944. That office had agreed to be responsible for validating the test and for conducting such additional statistical analysis of the data as seemed desirable and useful.

The pressure of other duties prevented the Personnel Research Section, AGO, from completing this work. It remained, at the end of the war, in the form in which the Applied Psychology Panel turned it over to the Army. The individual items are promising in that they have no easily picked out "right" answers. But considerable further work will eventually be required to determine how and under what conditions it can be best used to aid in the classification of Service personnel.

[redacted]

Chapter 6

THE SELECTION OF RADIO CODE OPERATORS

By *Dael Wolfle*[a]

Summary

At navy request the Applied Psychology Panel developed an improved test of radio code aptitude. The test first requires the subject to learn three simple characters of the radio code and then measures his ability to differentiate these characters as they are sent at faster and faster speeds.

The test was demonstrated to have a validity of .50 and to be superior to all other available code aptitude tests. It was officially adopted by both Army and Navy for use in the selection of men for radio code training.

6.1 INTRODUCTION

The process of acquiring skill in the reception of International Morse Code can be broken down into two phases. In the first phase the student learns 36 characters representing the alphabet and the digits. This task is relatively simple and ordinarily requires only a few hours or at most a few days of practice. In the second phase the student must learn to respond correctly to the characters when they are sent more and more rapidly. Ordinarily from 10 to 20 weeks of practice are required before the student acquires enough speed and skill to become a useful operator.

Some men fail to learn the individual characters satisfactorily, and many who do learn them fail to acquire sufficient speed to be useful as military or commercial operators. The result is a high attrition rate unless the students are very carefully preselected for code learning ability. Failure rates as high as 40 to 50 per cent were found in some schools early in World War II.

The problem of selecting code operators became apparent in World War I, and efforts were made at that time to develop suitable tests for that purpose. One of the tests developed then became standardized as the Signal Corps Code Aptitude Test (SCCAT) early in the nineteen-twenties. It remained as the standard selection test until after World War II began. Its low validity made the need for a better test obvious.

When the Committee on Service Personnel—Selection and Training was organized in June 1942, one of its early requests was to establish a research project on the selection and training of radio code operators for the Navy. The directive specifically ordered that research should be conducted to improve the techniques of selecting personnel for assignment to radio code instruction. The request was accepted and assigned to the Psychological Corporation, as the contractor, as NDRC Project N-107. Attempts to build better code tests were being made at the same time by other groups, both military and civilian.

6.2 TYPES OF CODE APTITUDE TESTS

Three chief types of code aptitude tests have been tried.

1. *Discrimination tests* measure the ability of the subject to distinguish between complex rhythmic patterns of dots and dashes. SCCAT is the best known example.
2. *Code learning tests* measure the ability of the subject to learn code characters. The test developed by L. L. Thurstone and an Army test (ROA-2) are examples.
3. *Speed of response tests* measure the ability to copy a few easily learned code characters at rapid rates of transmission. The NDRC Speed of Response (SOR) test is the best example.

The NDRC project worked with all three types. Preliminary results showed the third to be most promising, so work was concentrated on the development of a speed of response test.

[a] This chapter is based primarily upon the work of NDRC Project N-107.

6.3 THE NDRC SPEED OF RESPONSE TEST OF CODE APTITUDE[4]

6.3.1 Development

The idea for this type of test came from Biegel.[6] After introducing his students to the radio code characters, he presented signals to them in groups of 30, with less and less space between the individual characters in the successive groups. His test was long, apparently consisting of 900 characters in addition to the introductory learning series. Further, it was administered three times in order to increase its reliability. His data indicated that it appeared to have a reasonably high value in predicting subsequent achievement in radio code classes.

A speed of response type of test has two parts. First, there is a *learning unit* in which the subjects are taught a few characters. There have been several opinions as to the degree of difficulty of the letters to be taught. NDRC Project N-107 finally chose three very easy characters on the principle that it was desirable to have nearly all the subjects able to copy the characters at a slow speed of presentation, so that the test itself would be almost entirely of the ability to respond quickly to these signals. Second, there is a *testing unit* in which the subjects are tested on these learned characters at increasingly faster rates of presentation.

In the first edition the letters D (– . .), S (. . .) and R (. . .) were used. After these three letters were learned, the test itself began. It consisted of the three letters, D, S, and R, sent in random order, first at slow speeds and then more and more rapidly until the fastest speed was reached. It was then repeated in reverse order going from the faster down to the slower speeds, with the slowest speed omitted.

This initial form was found to have considerable predictive ability, but it contained a number of faults. The test was too long. The initial part, however, was too short; some of the students failed to learn the three characters well enough to distinguish them properly even at the slowest speed.

Form 2 attempted to correct these faults by using three easier letters, I (. .), N (– .), and T (–), by increasing the amount of learning drill, and by shortening the test itself.

The test went through two other preliminary editions in which refinements were made and in which it was adapted to machine scoring. Final editions were prepared, one for the Army and one for the Navy, a manual of instructions was written,[3] and the SOR was put into routine use for selecting radio operators for military duty.

6.3.2 Validity

Several criteria can be used to determine the validity of a code aptitude test. The final grade assigned in code school usually depends upon several factors, only one of which is code receiving speed. Even so it has been used in some studies, as has the dichotomous criterion of failing or passing the course. Code receiving speed as determined by school examinations has also been used. The chief difficulty with this criterion is that both test content and admin-

TABLE 1. Validity of Speed of Response Test.

Study conducted by	Criterion used	Weeks of instruction	No. of classes	No. of men	Median validity coefficient
NDRC	CRT	5	2	216	.30
NDRC	CRT	8	10	1,932	.34
Army	CRT	7–8	4	178	.41
Army	Code speed	7–8	4	221	.55
Navy	Code speed	14	1	185	.37
Navy	Pass-fail*	14	1	299	.49
Thurstone	Code speed*	6	1	215	.61
Thurstone	Pass-fail*	16	1	196	.57

*Biserial correlation; others were Pearson correlations.

[redacted]

istration procedures vary from school to school. In order to provide a standardized criterion, Project N-107 developed the Code Receiving Test (CRT).[1,2] This test consisted of phonographically recorded instructions, practice exercises, and test content. The test content consisted of five-character groups in which the 36 characters each appeared with equal frequency. Letters and numbers were presented in a purely random order, so they frequently occurred in the same group. CRT was therefore more difficult than the usual school test of receiving speed; a speed of 14 groups per minute being the equivalent of 20 words per minute of plain language material. The entire set of CRT records included tests at 12 speeds varying from 4 to 25 groups per minute.

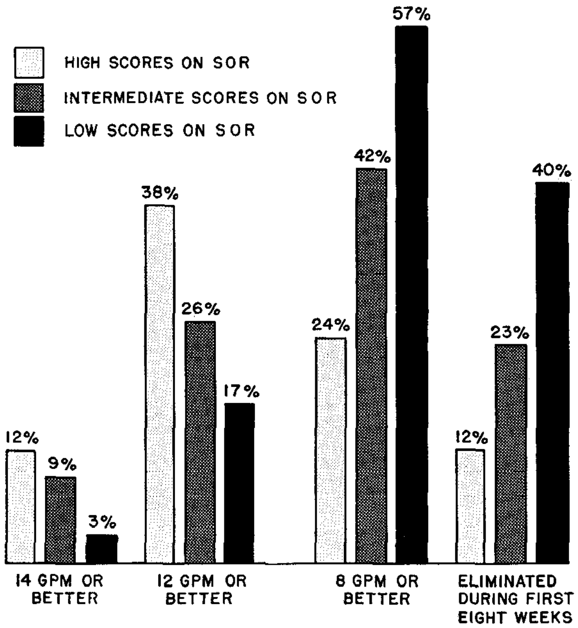


FIGURE 1. Relation between scores on Speed of Response Test and performance in radio code school.

Men were first classified as high (10 to 15 per cent of sample), intermediate (10 to 15 per cent), or low (70 to 80 per cent) in terms of score on SOR. The vertical bars show the percentage of each group having code speeds, after eight weeks of training, shown at the bottom of the figure. For example, 12 per cent of the high-score group, but only 3 per cent of the low-score group, had achieved speeds of 14 groups per minute or higher. $N = 1,001$.

The validity of SOR was determined by several investigators, using different criteria, on different groups of subjects, and after different amounts of code instruction. Eight of the resulting validity coefficients are reproduced in Table 1. They range from .30 to .61, with half of them .49 or higher. The validity shown by these correlations was considerably higher than that of the Signal Corps Code Aptitude Test which had been the standard selection test for approximately 20 years.

Validity of the SOR is shown in terms of practical usefulness in predicting code school success in Figures 1 and 2. Figure 1 analyzes the records of 1,001 men after eight weeks of training. The men in the class who had achieved speeds of 14 groups per minute or better, those who had achieved 12 groups per minute or better, those who had not yet gotten better than 8 groups per minute, and those who had

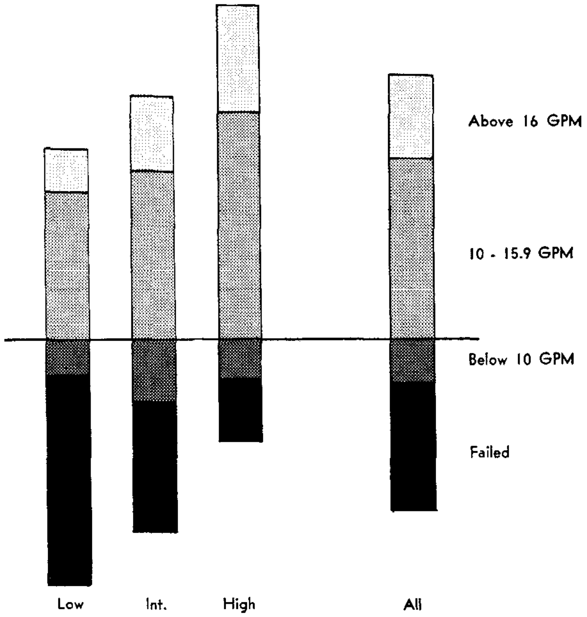


FIGURE 2. Speed of Response Test scores and code speed after 12 weeks of training.

The distribution of code speeds attained in 12 weeks of training is shown for three groups of men classified according to their SOR scores and for the total group of 616 men.
Low—bottom 70 to 80 per cent in SOR scores
Int.—next 10 to 15 per cent in SOR scores
High—top 10 to 15 per cent in SOR scores
All—the three groups combined

already been eliminated from code school are shown, divided according to whether their SOR test scores were high, intermediate, or low. Validity of the test is shown by the fact that

[redacted]

men who made high SOR scores more frequently attained high speeds, and were less frequently found in the poor groups, than men whose SOR scores were intermediate or low.

In Figure 2, results are shown in a different way for 616 men after 12 weeks of code training. Again, the men making high SOR scores less frequently failed and more frequently attained higher speeds than did men making low or intermediate SOR scores.

6.3.3 Comparison with Other Tests

The evidence summarized in Section 6.3.2 demonstrated the validity of the SOR. But other code tests were also available, and the Army and Navy wanted to adopt the most satisfactory one. Each Service, therefore, conducted an experimental comparison of several possible tests.

The Army comparison[8] of six tests, using two criteria, was based on a total of 221 men in four classes at Scott Field. The results are summarized in Table 2. The SOR was found to have a slightly higher validity than the Thurstone[7] test. Both were considerably better than any of the Army radio code tests or the Army General Classification Test in selecting good operators.

TABLE 2. Army comparison of six tests for predicting success in radio code training. ($N = 221$)

Test	Correlation with School speed	Correlation with CRT speed
SOR	.55	.41
Thurstone's test	.47	.44
ROA, X-1	.33	.39
SCCAT, I	.30	..
SCCAT, II	.33	..
AGCT	.24	..

The Navy comparison[9] of six tests, using three criteria, was based on samples of 183 to 306 men in the radioman school at Bainbridge Naval Training Center. The results are summarized in Table 3. The SOR and Thurstone tests were again found to have the highest validity coefficients. The other four tests (Clerical Aptitude, Spelling, Reading, and GCT), were all considerably less efficient than the Thurstone and NDRC code aptitude tests.

TABLE 3. Navy comparison of six tests for predicting success in radio code training.

Test	Correlation with 14th week code speed	Correlation with final average course grade	Correlation with Pass-fail in course
SOR	.37	.28	.49
Thurstone's test	.36	.31	.42
Clerical aptitude	.06	.14	.26
Spelling	.21	.22	.28
GCT	.12	.15	.09
Reading	.19	.18	.16

The superiority of SOR over Thurstone's[7] test, as shown in Tables 2 and 3, was slight. It was, however, a shorter test. The playing time of the records was 27 minutes for SOR and 44 minutes for Thurstone's test. Both tests had the necessary characteristics of being simple to administer, suitable for group administration, and designed for scoring with the International Business Machines Company scoring machine.

6.3.4 Adoption by Army and Navy

The Army and Navy tests[8, 9] referred to in Section 6.3.3 both led to official adoption in 1944 of the SOR for routine Service use. The Coast Guard and Maritime Service also used it.

Special recordings of the test were made by the project for Army and Navy use.

The Navy edition was recorded on two sides of one record to be played at 33⅓ rpm. It is identified as *Radio Code Test: Speed of Response, Form 2,* Columbia Recording Corporation serial number YTNY 2902-2903. The answer sheet is NavPers 16573.

The Army edition was recorded on eight sides (four records) to be played at 78 rpm. The test content is identical with the Navy form; minor differences in instructions were made to adapt it to Army use. The Army renamed the test *Army Radio Code Aptitude Test,* 1944, *ARC-1.* The records bear Columbia Recording Corporation serial numbers from XP33522 to XP33529. The answer sheet is IBM Form I.T.S. 1100 B 1133.

[redacted]

6.3.5 Advantages for Military Classification

The Navy data reported in Table 3 showed low correlations of code proficiency with the General Classification Test and tests of spelling, reading, and clerical ability. Age was also shown to have a low correlation with achievement in code school; for nine separate classes of from 123 to 258 men each, the median correlation with age was $-.14$ (range $-.22$ to $+.03$). Similarly, low correlations were found between code proficiency and education, mechanical ability tests, and tests of subject matter knowledge.[4]

Since youth, intelligence, education, and mechanical understanding are at a great premium for many military assignments, it is highly advantageous to have a test which predicts success in code school satisfactorily but does not correlate highly with these other variables. This situation permits both a negative and a positive type of selection for code training. Men with superior qualifications for other assignments can be eliminated from consideration, and selection then made from the remainder. By this means competent radio operators can be secured without taking men who are especially desirable for other assignments.

6.4 SELECTION FOR OSCILLOSCOPE CODE RECEPTION

In 1944 the Naval Research Laboratory requested Applied Psychology Panel assistance on problems of training men to receive code signals presented visually on an oscilloscope screen. In the course of training a number of men to receive such code, information was secured on the relation between SOR scores and learning records.[5] In a group of 57 men a biserial correlation of .50 was found with a criterion of meeting or failing to meet a specified proficiency level after three weeks of training. Rank order coefficients of .57 and .52 were found between SOR scores and proficiency at two code speeds. These correlations are as high as those obtained with auditory code.

A visual adaptation of the SOR should be tried out, even though it might not work better than the present auditory test. Present results indicate that the principle of the SOR test may be as useful in selecting men for visual code reception as for auditory code.

Chapter 7

SELECTING RADAR OPERATORS

By *Donald B. Lindsley* [a]

Summary

A NUMBER OF TESTS were designed for use in selecting and screening radar operators. The tests were intended to measure aptitude for the visual tasks of reading and interpreting oscilloscope patterns. Most of the tests showed satisfactory reliability. Validity data were unsatisfactory because of the lack of an adequate criterion of proficiency for radar operators.

Two of the tests were adopted by the Navy for use in the Combat Information Center Aptitude Test, Form 2. This test showed correlations of .45, .55, and .56 with final grades in three successive classes at naval training school (tactical radar).

7.1 INTRODUCTION

Despite the fact that radar played a critical part in helping win World War II and was from the start recognized as an important and specialized operation, little was done to establish standards of operating proficiency and to select operators in terms of such standards. The rapid development, production, and extension of field uses of radar meant that thousands of men had to be trained quickly. Training schools mushroomed almost overnight. These schools were usually overcrowded and understaffed; frequently there was insufficient equipment for proper teaching and demonstration; and in order to meet quotas, training was often extremely brief. Although a selection problem existed continually, few selection standards were developed and these were seldom rigidly applied due to the continued demand for operators. This demand prevailed throughout World War II since new types of radar meant new trainees or retraining of previously trained operators. Toward the end, when the demand for operators abated slightly, the number of men being inducted was smaller and the quality poorer; consequently the number available for training was reduced and it was still difficult to apply rigorous selection standards.

The principal criteria of Army and Navy specifications for radar operators were as follows: *age,* under thirty; *intelligence,* slightly above average; *vision,* average or corrected to average; and *interest* or experience in related fields. The establishment of these criteria was arbitrary.

Early in 1943 the Army requested a research project on psychological factors in radar operation. A Navy request for similar studies was made shortly afterward. The work was assigned, as Project SC-70, NS-146, to the Yerkes Laboratory of Primate Biology (Yale University). One of the first tasks undertaken was to construct a series of tests for the purpose of selecting and screening radar operators. Requests from Army and Navy training centers called for screening measures to eliminate men already assigned for radar training. It was hoped that such screening would result in meeting quotas with more proficient operators and also save valuable training time and facilities otherwise wasted on men who were unable to pass the course and meet minimal standards of proficiency. Not only were selection standards poor, but in many instances little attention was paid to final proficiency; the inept were often passed and took their places as relatively inefficient operators.

After a survey of the types of functions performed by radar operators, it was concluded that one of the critical aspects of the radar operator's task was visual perception, especially the ability to detect visual changes quickly and accurately, and to note their movement, form, and spatial relationships. Sustained attention and emotional stability under pressure also were recognized as important factors.

A series of pencil-and-paper tests, believed to embody measurements of some of these ca-

[a] This chapter is based upon the work of Project SC-70, NS-146.

pacities, was constructed. These tests were then tried out in Army and Navy training centers in order to determine their reliability and validity. Although reliability was readily ascertained, validity presented a problem which was not satisfactorily solved. Validity required a suitable criterion of proficiency against which to correlate the test results. The principal criteria of proficiency available at the time were school grades and instructor ratings of performance during training. Neither of these proved to be reliable measures. The examinations upon which course grades were based frequently were weighted heavily with nonessentials and seldom were comparable in difficulty from class to class or from school to school. Ratings of performance presented similar difficulties. Later it was possible to demonstrate that objective ratings of performance could be devised and that proficiency measures could be developed which were objective and discriminative.

In retrospect it now appears that it would have been profitable to have concentrated initially on the development of proficiency measures to be used at the termination of training, in order that a satisfactory criterion of performance might have been available for correlation with test results.

7.2 DEVELOPMENT OF SELECTION TESTS FOR RADAR OPERATORS

In May 1943, Project SC-70, NS-146, prepared a series of thirteen pencil-and-paper Oscilloscope Operator Tests.[1,2] The tests measured perceptual and speed capacities thought to be essential to radar operation. Table 1 names the tests and summarizes the chief data available concerning each. The tests are printed in reference 2. Some of the tests, especially Nos. 4, 5, 9, 10, 11, 12, and 13, simulated closely some of the tasks which a radar operator is required to perform and therefore had apparent "face validity." For example, the Ratio Estimation Test presented a series of items which required the estimation of the ratio of paired vertical lines of different height. This was a duplication of the task presented by the A-scope on a *ground control interception* [GCI] radar used for determining elevation of targets, where the ratio of two simultaneous pips had to be judged by the operator. The Coordinate Reading Test presented a simulated *plan position indicator* [PPI] scope face on which there were a large number of blips whose range and bearing were to be read. Similarly the other tests reproduced some aspects of the operator's task. Thus quantitative measures of speed and accuracy of performance could be obtained.

7.2.1 Reliabilities and Intercorrelations of the Tests

DREW FIELD STUDY

After preliminary tryout with enlisted men at Camp Murphy, Florida, 12 of the 13 tests were administered to 378 radar operators-in-training at Drew Field. The purpose of this study[3] was to determine the reliability and validity of the tests as selection or screening instruments. Unfortunately no satisfactory validation data could be obtained in this study since the criterion of operator proficiency which consisted of school grades and ratings proved to be unreliable statistically and therefore inadequate for correlation with the test results. Distributions of scores for most of the tests were found to give adequate discrimination. The reliability and intercorrelations of the tests, based on a sample of 100 randomly selected men, were determined. These data are presented in Table 1.

The reliability of all tests was satisfactory, except for the scale reading and oscilloscope reading tests, which were subsequently revised and lengthened. Except in a few instances sufficient independence of measurement was found to warrant further experimental study of the individual tests with the view to establishment of predictive batteries. The correlations of the tests with the Army General Classification Test (AGCT) were all positive and ranged from .19 to .63. It was recommended that the AGCT be included in any battery of tests for the selection of radar operators.

The tests had adequate reliability for individual prediction. Although intercorrelations indicated that the tests dealt with related func-

[redacted]

tions, it was decided that two or more of them might well be included in the same selection battery. It was further decided to try the tests out in another situation in which more adequate validating data might be obtained. Accordingly, the following study was made.

Orlando Study

An adequate criterion of radar operator performance was believed to be available in the eight operating stations of the Army Air Forces Tactical Center, Orlando, Florida.[7] Here men of varying degrees of experience in operational training were performing operating duties daily under the supervision of experienced instructors. Two officers in charge at each station rated the men on operating performance. Ratings were made on A-scope reading, plotting ability, PPI scope reading and GCI operation. The reliabilities of the ratings in five of the eight stations were found to be satisfactory. Accordingly, a study of the relationship between the test results and the ratings of operator proficiency in these five stations was made. The different stations served *Early Warning* (EW), *Chain Home Link* (CHL) and *Ground Control Interception* (GCI) functions. Correlations were computed for the separate stations; the number of men in individual stations ranged from 14 to 36.

Most of the tests gave promising validity coefficients; the six providing the best prediction possibilities were as follows: Scale Reading, Course Location, Plot Reading, Ratio Estimation, Coordinate Reading, and Polar-Grid Coordinate. The validity coefficients for the first four of these tests, in the individual stations, ranged from .35 to .78 with 17 of the 19 coefficients above .50 and 9 above .60. Coordinate Reading, measuring proficiency in PPI reading, and Ratio Estimation, measuring ability to estimate pip ratios and to match pips, were shown to distinguish between different levels of training and experience. The Plot Reading, Scale Reading, and Polar-Grid Coordinate tests were particularly applicable as selection devices for predicting success in various types of radar operation.

TABLE 1. Information concerning Oscilloscope Operator Tests.[2, 3, 7, 9]

Name of test	Face validity	Reliability (split-half, corrected)	Mean correlation with other tests	Correlation with AGCT	Mean validity RHO[7]	Used by Navy†
1. Form Detection		.96	.33	.22		
2. Form Conversion I		.97	.43	.39		
3. Form Conversion II		.92	.37	.45		
4. Scale Reading	x	.81	.45	.63	.60	CIC
5. Oscilloscope Reading	x	.86	.40	.30		
6. Course Location					.74	
7. Spot Location		.98	.42	.20		
8. Target-Course Analysis		.96	.27	.19		
9. Plot Reading	x	.94	.42	.42	.65	
10. Ratio Estimation	x	.96	.39	.47	*	TR
11. Coordinate Reading	x	.90	.47	.45	*	TR
12. Coordinate Plotting	x	.94	.46	.43		
13. Polar-Grid Coordinate	x	.90	.51	.47	.53	TR, CIC

*Discriminated significantly among operators with different amounts of training.
†TR = Tactical Radar Aptitude Test (NavPers 16574).
CIC = CIC (Combat Information Center) Aptitude Test, Form 2 (NavPers 16980).

Other Results

Two of the above tests, Polar-Grid Coordinate and Scale Reading, were given at the Naval Training School, Virginia Beach, Virginia, to 120 search radar operators. The principal validating criterion was the final course grade. The correlations between test results and final grades in the radar course were .50 and .55 respectively.

During the summer of 1943, three additional tests were constructed. These were the Airborne

[redacted]

Oscilloscope Reading Test, Oscilloscope Conversion Test, and Oscilloscope Interpretation Test. The first two are concerned with airborne radar operator functions, especially interpretation and transmission of symbolic directions from operator to pilot. Both tests were tried out on AI (airborne interception) operators at Boca Raton Field and were found to be not highly predictive of success as measured by school grades. The Oscilloscope Interpretation Test is a test of ability to detect different types of radar signals through varying degrees of background masking or jamming. This test was administered at the Naval Training School, Fort Lauderdale, Florida. Reliability was satisfactory, but validation data could not be obtained. The test was tried out by the Services in an AAF navigation school for selecting radar operator trainees and also in a special research project in the Eighth Air Force.

7.2.2 Application of the Tests by the Services

ARMY

In May 1944, upon request from the Psychological Research Section of the Air Surgeon's Office, copies of the following tests were supplied for experimental tryout in selecting pathfinder crews in the Eighth Air Force: Target-Course Analysis, Coordinate Plotting, Form Conversion I, Form Conversion II, Oscilloscope Reading, and Airborne Oscilloscope Reading tests. The results of these studies are not known.

In the summer of 1944, upon request from the Air Surgeon's Office, special machine-scored editions of the Scale Reading Test and the Oscilloscope Interpretation Test were made available for use at Selman Field, Louisiana, in the selection of radar operator trainees. The Polar-Grid Test was revised by representatives of this station for use in connection with the same problem.

The Coordinate Reading Test and the Oscilloscope Interpretation Test were tried out in a battery of tests for the selection of bombardier instructors at Midland Field, Texas. It is understood that the tests gave sufficient promise to be incorporated in a radar operator selection battery.

NAVY

Tactical Radar Aptitude Test.[8, 9] Three of the tests, the Polar-Grid Coordinate Test, Ratio Estimation Test, and Coordinate Reading Test were included in the Tactical Radar Aptitude Test (NavPers 16574) which was included as part of the selection requirements established for naval training school (tactical radar).

The Polar-Grid Coordinate Test measures the examinee's ability to translate the reading of a point on a polar coordinate to a grid coordinate chart. The Ratio Estimation Test measures the ability to estimate the relative lengths of lines presented in pairs. The Coordinate Reading Test measures the ability to estimate the direction and range of targets on a polar coordinate chart.

The Bureau of Naval Personnel tried out the Tactical Radar Aptitude Test on several classes of approximately 100 men each.[9] Using several criteria of success in Tactical Radar School, validity coefficients were generally small, ranging from —.19 to +.34. Twenty-seven of a total of 40 correlations were .20 or lower.

The CIC Aptitude Test. On the basis of these results, however, the Bureau of Naval Personnel constructed a new test, the CIC Aptitude Test, Form 2 (NavPers 16980).

This test retained the Polar-Grid Coordinate Test from the Tactical Radar Aptitude Test but dropped the other two. In their places were included the Scale Reading Test and a Relative Movement Test which was originally developed by the University of California Division of War Research under Section 6.1 of NDRC.

At the same time that the selection tests were being revised and improved, work was under way in improving the examinations and final grades given at NTSch (tactical radar). As a result of both of these improvements, the CIC Aptitude Test Form 2 gave much higher, and much more consistent, prediction of school success than had the Tactical Radar Aptitude Test. Detailed results are shown in Table 2. Correlations with final grades in three successive classes were .45, .55, and .56. These results

demonstrated the usefulness of the test as a selection device. A minimum Navy standard score of 50 was, therefore, included as one of the selection requirements to be observed in the selection of candidates for tactical radar training.

7.3 RELATIONS BETWEEN VISUAL PROFICIENCY AND RADAR OPERATION

The following study was made to determine whether there was a relationship between certain visual capacities as measured by the Bausch and Lomb Ortho-Rater and proficiency in radar operation. This study[6] was carried out in field operating stations of the Army Air Forces School of Applied Tactics, Orlando, Florida. Visual measurements, including binocular and monocular visual acuity (near and far), vertical and lateral muscle balance (phoria), stereopsis, and color vision, were obtained for 157 radar operators. Ratings of proficiency in scope operation were also obtained. In general it was found that operators with *substandard binocular acuity at near distances* were rated as less proficient than those with normal or better acuity; also that operators with *excessive overconvergence at near distances* were rated lower than those with normal convergence. In view of these results, tests of binocular acuity (near) and lateral phoria (near) were recommended for examination of prospective radar operators. It was recommended that minimum standards be set at 1.0 (decimal notation) for the near acuity test and at 6 prism diopters of overconvergence or esophoria.

7.4 USE OF PROFICIENCY MEASURES AND RADAR TRAINERS AS SCREENING DEVICES

The usual procedure is for men to proceed from "boot camp" or a classification center to a training center for primary training in an assigned specialty. In the case of radar, therefore, it would be highly desirable to have selection standards which could be applied before a man is assigned to training, in order to eliminate those who would probably be "washed out" in training.

There is, however, another point at which screening and selection may occur. This is after primary, but before advanced or operational, training. Also it frequently happens that small groups are selected from larger groups for specialized duty or performance which calls for exceptional skill and ability. For these secondary types of selection the use of proficiency or achievement measures of an objective and performance type is recommended. Radar proficiency measures and their use in the classification and assignment of men are described in references 4 and 5. Such measures, particularly where knowledge of radar equipment and its functional uses was involved, have been shown to differentiate between expert, average, and apprentice levels of performance. Thus the proficiency measure may be used at the end of one stage of training to select those who will advance to the next stage.

Frequently a man will do well in courses where knowledge of equipment is required but will fail in operational performance. In so far as certain radar trainers provide good simulation of operating conditions and at the same

TABLE 2. Reliability and validity of CIC Aptitude Test, Form 2.

		Correlations with CIC achievement test			Correlations with final grades		
	Reliability	Class 7L	Class 8L	Class 9L	Class 7L	Class 8L	Class 9L
Polar-Grid Coordinate	.85	.49	.37	.42	.39	.45	.43
Scale Reading	.85	.55	38	.56	.46	.51	.56
Relative Movement	.82	.48	40	.47	.31	.51	.39
Total test	.92	.61	.45	.57	.45	.55	.56

time make possible quantitative measures of performance, which are often impossible to secure with the actual equipment, they are valuable for assessing some of the performance aspects of radar operation. The measures obtained may be used to eliminate men who would not profit from further training. This implies, of course, that a significant relationship exists between performance on the trainer and operational performance with the actual gear. During the war there were a few radar trainers which seemed to meet requirements for use as selection devices. It should be emphasized also that, if the relationship between trainer performance and actual radar operation is high enough, the trainer may be used as a validating criterion against which to correlate the results of pencil-and-paper radar aptitude tests. This is particularly important since one of the chief obstacles in the development of radar aptitude tests has been in finding suitable and adequate criteria of proficiency of radar operation.

Finally, if the validation of radar aptitude tests must rest on success or proficiency attained in various stages of training rather than upon final operational proficiency, which may be impractical to measure, it should be emphasized that effort should be placed initially upon the improvement of proficiency or achievement measures. These must be made comprehensive and objective; performance functions should be emphasized and should be given a weighting proportional to their value in the final assessment of the operator. Without a reliable and adequate measure of proficiency it is impossible to establish the validity of a selection test.

7.5 FUTURE DEVELOPMENT OF RADAR APTITUDE TESTS

There are a variety of radar applications for land, sea, and air operations, each type of equipment differing in its uses, characteristics, and operating procedures. It is probable that operator requirements differ for each application and that selection standards will have to vary accordingly. During the course of World War II, minimal use of radar aptitude selection devices occurred and comparatively little progress was made in developing and introducing such selection measures. The main pitfall seemed to be the problem of finding an adequate validating criterion. The solution to this problem lies in working out a plan for utilizing operational criteria, either from overall field performance or from specialized, objective methods of evaluating operational performance during training. The improvement and objectification of proficiency measures, such as achievement examinations, ratings of performance, or radar trainer scores, is a necessary prerequisite to the development of better selection tests.

Chapter 8

SELECTING STEREOSCOPIC RANGEFINDER AND HEIGHTFINDER OPERATORS

By *William E. Kappauf, Jr.*[a]

Summary

EARLY IN ITS HISTORY NDRC undertook work on the improvement of stereoscopic heightfinders. Efforts to improve this equipment led quickly to the development of related efforts to improve the selection and training of heightfinder operators. This psychological work was directed at first by Division 7. When a specialized psychological section was established in June 1942 (first as the National Research Council Committee on Service Personnel —Selection and Training, and later as the Applied Psychology Panel, NDRC), responsibility for the work was transferred to it.

Standards for the selection of stereoscopic heightfinder operators were developed, recommended to and adopted by the Army, and validated by the project. Assistance was given to the Army in setting up stereoscopic testing centers where men to be trained as heightfinder operators were selected.

When the Navy established a school for training men for the rate of fire controlman (R), the Army heightfinder standards were used as the basis for setting selection requirements for rangefinder operators. The Applied Psychology Panel work was extended at the same time to include studies of the selection and training of rangefinder operators. Improved selection standards were recommended to the Navy and adopted for general use. The project staff trained Navy personnel for duty in the two testing centers where men to be trained as rangefinder operators were selected.

In the course of the work on improving selection standards, information was obtained on the reliability of a number of visual test instruments: the Shuron pupillometer, the NDRC interpupillometer, the modified Massachusetts vision test kit, the Ortho-Rater, the projection eikonometer, the stereoscopic trainer M2, the vectograph-pursuit test, and the Dearborn-Johnston test.

Correlations among the several tests of stereoscopic vision were low enough to indicate that these tests were not measuring the same kind of ability and could not be used as substitutes for each other. A thorough study of military visual requirements and of improved methods of visual testing is recommended.

[a] This chapter is based on work done by the Princeton University Fire Control Research Project of NDRC Division 7, and by Project N-114 and the Heightfinder Project of the Applied Psychology Panel. A supplementary report is contained in Volume 2 of the Summary Technical Report of Division 7.

8.1 PROBLEMS IN VISUAL SELECTION

The obvious importance of good vision in a wide variety of military situations and tasks led to a great deal of work on visual tests and visual standards. The Applied Psychology Panel was active in part of this work, dealing in particular with problems of stereoscopic vision and night vision.

Very little is known regarding the actual visual requirements of most jobs, or the minimum visual standards which should be imposed in selecting men for a particular job. The practice, in view of this ignorance, has often been to set very high visual requirements for any military job in which vision appears to be important. When this is done for many jobs, it soon becomes difficult to find enough men who meet the standards. The result is that requirements are then completely disregarded or are modified by classification personnel who are usually unfamiliar with the relative importance of the several standards. To be reasonable, selection standards should set the minimum requirements which will keep school or training failures down to a tolerable number but not

impose impossible requirements on a classification department.

The Applied Psychology Panel was asked to investigate the severity of the selection standards for stereoscopic rangefinder and heightfinder operators. The work led to the development and improvement of some of the testing equipment, to validation experiments, and to some developments in instrument maintenance and operator training to ensure dependable validation test results.

8.2 HISTORICAL SUMMARY OF WORK ON RANGEFINDER AND HEIGHTFINDER SELECTION TESTS

Qualifications for stereoscopic heightfinder operators were listed in the Army heightfinder pamphlet in use before the war. These qualifications included 20/20 vision or better in each eye, hyperphoria less than ½ prism diopter, esophoria less than 12 prism diopters or exophoria less than 6 prism diopters, interpupillary distance between 58 and 72 mm, general good health, and no history of eyestrain or fatigue. These qualifications had been established on an a priori, not an empirical, basis. Their application rested entirely with the battery commanders at the time that men were to be detached for training at the heightfinder school. Some of the men who arrived at the school met the standards; others did not, as indicated by visual acuity and other tests given them at the school. Interviews revealed that many men assigned to the school had no interest in heightfinder work. Selection was certainly not the best, and this coupled with lack of student interest meant mediocre heightfinder operation.

8.2.1 Studies at Fort Monroe, Virginia

When the Princeton University Fire Control Research Project was set up under Division 7 early in 1941, one of its objectives was to formulate a satisfactory selection program for heightfinder operators. The new tests were to be tests of demonstrated merit. Accordingly, a wide variety of visual, physiological, and pencil-and-paper tests were given to each class of students at the school, and the relation between these test scores and final school grades was determined. By the end of the year, a good number of the original tests had been eliminated from further consideration in the test battery. Validity was indicated for certain tests of stereoscopic vision (notably three tests on an instrument called the projection eikonometer and one on the stereoscopic trainer M2) and for pencil-and-paper tests of intelligence and mechanical aptitude.[27] The task of obtaining satisfactory validation data had been difficult because the total number of students who had passed through the heightfinder school was less than 150 and they, through such selection as was applied, were more homogeneous than the general run of enlisted personnel. Finally in February 1942, Division 7 proposed a tentative new selection battery.[1]

Suggested minimum qualifications for stereoscopic heightfinder or rangefinder operators were:

General intelligence—an Army standard score[b] of at least 100 on the AGCT, or at least 85 on the Navy O'Rourke GCT. (These were erroneously believed to be equivalent scores.)

Mechanical comprehension—a grade of I, II, or III on Part 3 of the Army test MA2 or MA3, or at least 75 on the Navy O'Rourke Mechanical Comprehension Test.

Height—for Army only—not less than 5 feet, 6 inches (imposed by tripod height for Army heightfinders).

Interpupillary distance—60 to 70 mm (imposed by limits of instrument adjustment).

Vision—visual acuity of at least 20/20 in each eye—hyperphoria not greater than ½ prism diopter—exophoria not greater than 6 prism diopters—esophoria not greater than 6 prism diopters.

Stereoscopic vision—an Army standard score of at least 110 ($\frac{1}{2}\sigma$ above mean) on the projection eikonometer, and a standard score of at least 110 on the stereoscopic trainer M2 modified to include power driven change of range.

Desire for training—selecting only those men who indicated that they wanted training as a stereoscopic observer after the nature of an observer's duties had been explained.

---

[b] Army standard scores have a mean of 100 and a standard deviation of 20.

It was further proposed that special personnel be trained to administer these tests at selection centers. The results of the testing for each man were to be recorded on his qualification card.

With the cooperation of the Coast Artillery Corps, the proposed tests were administered for standardization purposes at Fort Eustis, Virginia. Analysis of the records from these standardization experiments indicated where improved administration was required and where test simplification was possible. In particular it was found that the projection eikonometer test could be reduced from three different tests to one, the stereo-vertical test.[c]

So that the Services could extend the program, detailed descriptions of the test equipment and testing procedures were provided.[2, 3, 6] The equipment included the Shuron pupillometer for measuring interpupillary distance, a modified Massachusetts vision test kit for measuring visual acuity and phoria, a projection eikonometer as developed from an instrument used earlier at the Dartmouth Eye Institute (Section 8.3.7), and a modified stereoscopic trainer M2. In accordance with the original proposals of Division 7, three stereoscopic testing centers were established, at Fort Eustis, Camp Wallace, and Camp Callan.

A complete summary of the data which provided the basis for the Division 7 selection recommendations is found in reference 27. It is of interest to note what tests showed no useful or significant relation to heightfinder performance. They included a static test of stereoscopic vision, a series of pencil-and-paper tests other than those concerned with general intelligence and mechanical comprehension, and a wide variety of ocular and general physiological tests. Visual acuity and phoria requirements were retained in the selection battery specifically for their usefulness in reducing testing time by weeding out men who would have difficulty passing the stereoscopic tests. Validation data on acuity and phoria were meager because the heightfinder school students did not show significant departure from normal in these respects.

8.2.2 Studies at Camp Davis, North Carolina

There were two specific and recognized limitations in the conditions of the experiments which lay behind the Division 7 recommendations for the selection of stereoscopic operators. The men tested in the school and on whom the tests were validated were too homogeneous. The criterion score which was used as the measure of heightfinder operation was a variability score, not a score of absolute accuracy. Since only 4 per cent of entering selectees could pass all the established selection requirements, there was speculation as to the possibility of getting comparably efficient selection with lower selection cost by using other stereoscopic vision standards. A follow-up validation study was therefore requested by the Antiaircraft Artillery Command.

The second validation experiment was conducted by the Applied Psychology Panel project on the Selection and Training of Heightfinder Operators.[12] This project, accepting the adequacy of existing selection requirements other than the stereoscopic ones, proposed to determine the most satisfactory selection test(s) of stereoscopic vision. Through the cooperation of the Antiaircraft Artillery Command, a series of classes with men of all levels of stereoscopic ability were admitted to the Heightfinder School at Camp Davis. Through the further cooperation of the A.A. Schools at Camp Davis, and the Heightfinder School staff in particular, the school heightfinders were so maintained and reference measures of target position were so determined that it was possible to score the students not only on the variability of their performance but also on their absolute accuracy. Four tests of stereoscopic vision were administered to all students prior to enrollment at the school. Two of these were the projection

[c] Throughout this STR and in the original research literature this test is called the projection eikonometer test. The name, however, confuses the test with others used by the Dartmouth Eye Institute in the analysis of a visual condition known as *aniseikonia*. To avoid this confusion it is strongly recommended that in future Service use, the present test be called the stereo-vertical test.

[redacted]

eikonometer test and the stereoscopic trainer test which Division 7 had recommended. The others were tests which had been developed under a Division 7 laboratory contract with Harvard University. All students were men who had expressed an interest in learning to operate stereoscopic instruments.

The data of the validation study, based on a validation population of 92, indicated that the most efficient selection test was the projection eikonometer (see data in Section 8.4.4). Furthermore, if the passing score on this test was raised from 110 to 115, the stereoscopic trainer test could be eliminated from the test battery. Use of the projection eikonometer test with a passing score of 115 eliminated fewer men in the selection process and admitted a school poplation of which a greater portion could be expected to graduate.

The practical results of the research were these: a change in the stereoscopic vision test requirement in the selection battery for heightfinder operators, a small increase in the number of selectees who could pass all tests (from about 4 per cent to about 5½ per cent), and a decrease in the amount of testing time required in order to qualify or disqualify a man.

8.2.3 Studies at Fort Lauderdale, Florida

On the basis of the foregoing research, the Navy set up a list of selection requirements for men entering schools to be trained for the rate of fire controlman (R). This was a new rate established in June 1943, and was to be awarded to men proficient in the operation of rangefinder and radar equipment. The selection standards adopted at that time were as follows:

General intelligence—a Navy standard score[d] of 55 on the GCT.
Mechanical aptitude—a Navy standard score of 45 on the MAT.
Reading ability—a Navy standard score of 50 on the Reading Test.
Interpupillary distance—between 60 and 70 mm.
Vision—visual acuity 20/20 in each eye uncorrected.
—hyperphoria not greater than ½ prism diopter.
—exophoria not greater than 6 prism diopters.
—esophoria not greater than 6 prism diopters.
Personal qualifications—emotionally stable, calmness under stress, quick reactions, not over 30 years of age.

As a result of recommendations from the Applied Psychology Panel to the Bureau of Naval Personnel and the Bureau of Medicine and Surgery, these requirements were modified in April 1944 in three respects.

General intelligence—the passing score on the GCT was dropped to 50.
Vision—visual acuity of 20/22 or better was acceptable for acuity measurements made on the Ortho-Rater (test raw score of 9).
Stereoscopic vision—the requirement added was a Navy standard score of at least 58 ($\sigma = 10$) on the multiple projection eikonometer.

Having adopted these selection standards, the Navy set up two selection centers, one at Sampson and the other at Farragut, to classify men for training for fire controlman (R). The two testing centers were staffed by men trained in the administration of the selection test battery by members of Project N-114, the Applied Psychology Panel.

At the further request of the Navy, Project N-114 carried out an experimental investigation of the adequacy of this set of selection standards for the Fire Control School at Fort Lauderdale, Florida. Actually this research program began before the above selection requirements were in effect, and so the student groups which were tested and examined at the school were representative of the unselected Navy population. The project sought a test battery which would be efficient in selecting men who could pass the theory and study part of the course and also succeed in becoming good rangefinder operators. Compiled course grades on theory and operation and a variability score for rangefinding during the final examination were taken as the validating criteria. The criterion population for the rangefinder operator tests numbered 141 men.

The result of this investigation[22] was the development of two test batteries which were found to be equally effective in terms of picking good students. Battery I below required longer testing time per man but passed more men than did Battery II. The batteries proposed were:

[d] Navy standard scores have a mean of 50 and a standard deviation of 10.

[redacted]

	Selection Battery I	Selection Battery II
General intelligence	GCT 50	GCT 45
Arithmetic ability	AR 45	AR 55
Interpupillary distance	60-70 mm	60-70 mm
Visual acuity and phoria	Visual acuity score of at least 9 for each eye on the Ortho-Rater Hyperphoria less than ½ prism diopter (Ortho-Rater score 4 to 7) Esophoria less than 3½ prism diopters, or exophoria less than 4½ prism diopters (Ortho-Rater score 4 to 12)	
Stereoscopic vision	Multiple proj. eikonometer score of at least 58.	Ortho-Rater depth score of at least 5.
Personal qualifications	As in the original Navy requirements; must want to become rangefinder operator.	

These lists of qualifications were recommended to the Navy in March 1945. No action was taken on adopting them.

8.2.4 Other Research

In the course of the research just reviewed, a great deal of information was collected on the reliability of various visual and nonvisual tests and on the relation between different test scores. These data will be of interest to persons participating in future testing programs or test research. The remaining sections of this chapter therefore discuss these data along with accounts of the quantitative results of the validation experiments.

8.3 DESCRIPTION AND RELIABILITY OF TESTS USED

Before presenting the details of the validation and testing programs, the various tests used in the course of the studies by the Heightfinder Project and by Project N-114 will be described.

The visual tests included the Shuron pupillometer, the NDRC interpupillometer, the modified Massachusetts vision test kit, the Ortho-Rater industrial vision test, the projection eikonometer, the stereoscopic trainer M2, the vectograph-pursuit test, and the Dearborn-Johnston test. Navy pencil-and-paper tests were also investigated to determine their usefulness in selecting men for fire control courses.

8.3.1 The Shuron Pupillometer

This instrument for measuring interpupillary distance was used in all the early selection work. It was first adopted and recommended for selection use by Division 7.[2, 3] It consists of a frame with two sliding glass panels. On each piece of glass is a vertical hairline. When the subject's eyes assume parallel lines of regard and the examiner bisects each of the subject's pupils with one of the hairlines, the distance between the hairlines is the measured interpupillary distance. The median of three measurements with this instrument has an average error of somewhat less than 0.5 mm.

8.3.2 The NDRC Interpupillometer

The NDRC interpupillometer[17] was developed to meet the need for accurate interpupillary distance adjustment in rangefinder and heightfinder operation. It is described in detail in Chapter 22 of Volume 2, Applied Psychology Panel. It incorporates the principle of the Shuron pupillometer in an instrument where the subject makes his own slide settings. The median of three measurements taken with this instrument has an average error of less than 0.25 mm.[14]

The NDRC interpupillometer was used in the Navy selection program. These instruments were not available in quantity for use in training stations and aboard ship, so each man had his interpupillary distance accurately measured at the selection center. The measurement was recorded on his qualifications card, for use when and if he began training for the job of rangefinder operator.

8.3.3 Interpupillary Distance Measures in Selection

Selection on the basis of interpupillary distance is necessary because the range of adjustment of the interocular distance on rangefinders and heightfinders is limited. These limits

[redacted]

are 58 mm and 72 mm. Selection standards use the range from 60 to 70 mm to allow for errors of measurement and errors of rangefinder interpupillary scale calibration. Actually, not many men are eliminated on the basis of interpupillary measures. More Negroes than whites

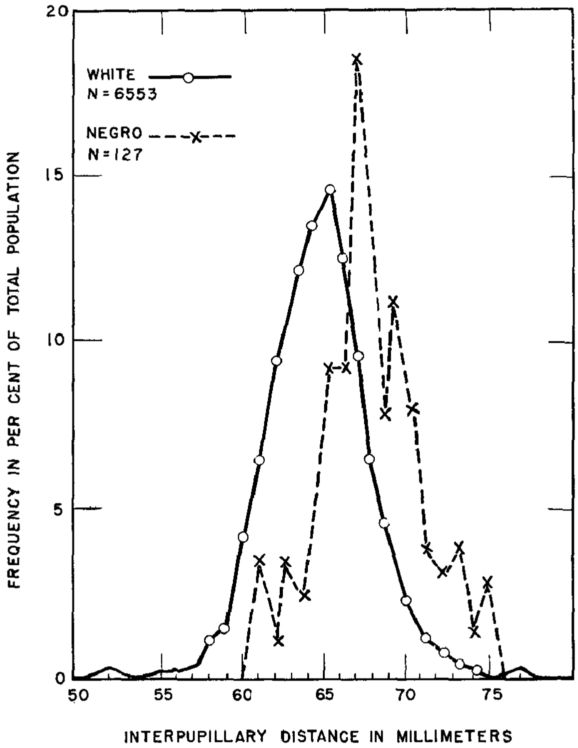


FIGURE 1. Distributions of interpupillary distances for white and Negro soldiers.

are eliminated on this basis, however.[9] Distributions of interpupillary distance measures for white and Negro populations are shown in Figure 1.

8.3.4 The Modified Massachusetts Vision Test Kit

The equipment in this kit consists of a visual acuity chart and phoria test materials including a chart and Maddox rod spectacles. The kit was developed and introduced by Division 7.[2, 4] The visual acuity chart uses Snellen E's. After the subject reads some larger letters, he is given 25 letters of the 20/20 size to read with his right eye and 25 to read with his left eye. Letter presentation is controlled by a curtain-slot arrangement which allows the subject to see only one row of letters at a time. The subject must read 80 per cent of the 25 letters correctly with each eye to qualify as having 20/20 vision. The reliability of this test at this 20/20 level is high, as indicated in the results of one experiment where 97 per cent of those who passed the test on its first administration also passed it on retest.[28]

Phoria measurements are taken by a Maddox rod technique. The subject views the phoria chart with his left eye. The chart has vertical and horizontal scales marked off in prism diopters. With his right eye, the subject sees a red streak, formed by the Maddox rod over his right eye and arising from a small electric light at the center of the chart. The position of the streak on the scales is the measure of phoria. The reliability of the test is comparable to that of other tests.[28] For $N = 152$, the test-retest correlation for vertical phoria was $r = +.59$ ($\sigma_r = .05$), and for lateral phoria was $r = +.83$ ($\sigma_r = .03$).

8.3.5 The Ortho-Rater Industrial Vision Tests

The Ortho-Rater is a visual testing instrument developed by the Bausch and Lomb Optical Company for use in industrial classification (see Figure 2). All tests on the instrument are given by viewing stereoscope-type slides. Among other measures of visual function which can be obtained with the Ortho-Rater are far vision tests of visual acuity (right eye, left eye, and binocular), phoria (vertical and lateral) and stereoscopic acuity. These tests were examined by Project N-114 for possible use in the selection of Navy rangefinder operators.

The reliability of the Ortho-Rater tests as administered to Naval personnel by Naval examiners was the first step in the investigation.[18] A test-retest study was carried out on a population of 234 recruits at Sampson. Results are given in Table 1. Retests were always given by a different examiner from the one who gave the original test.

[redacted]

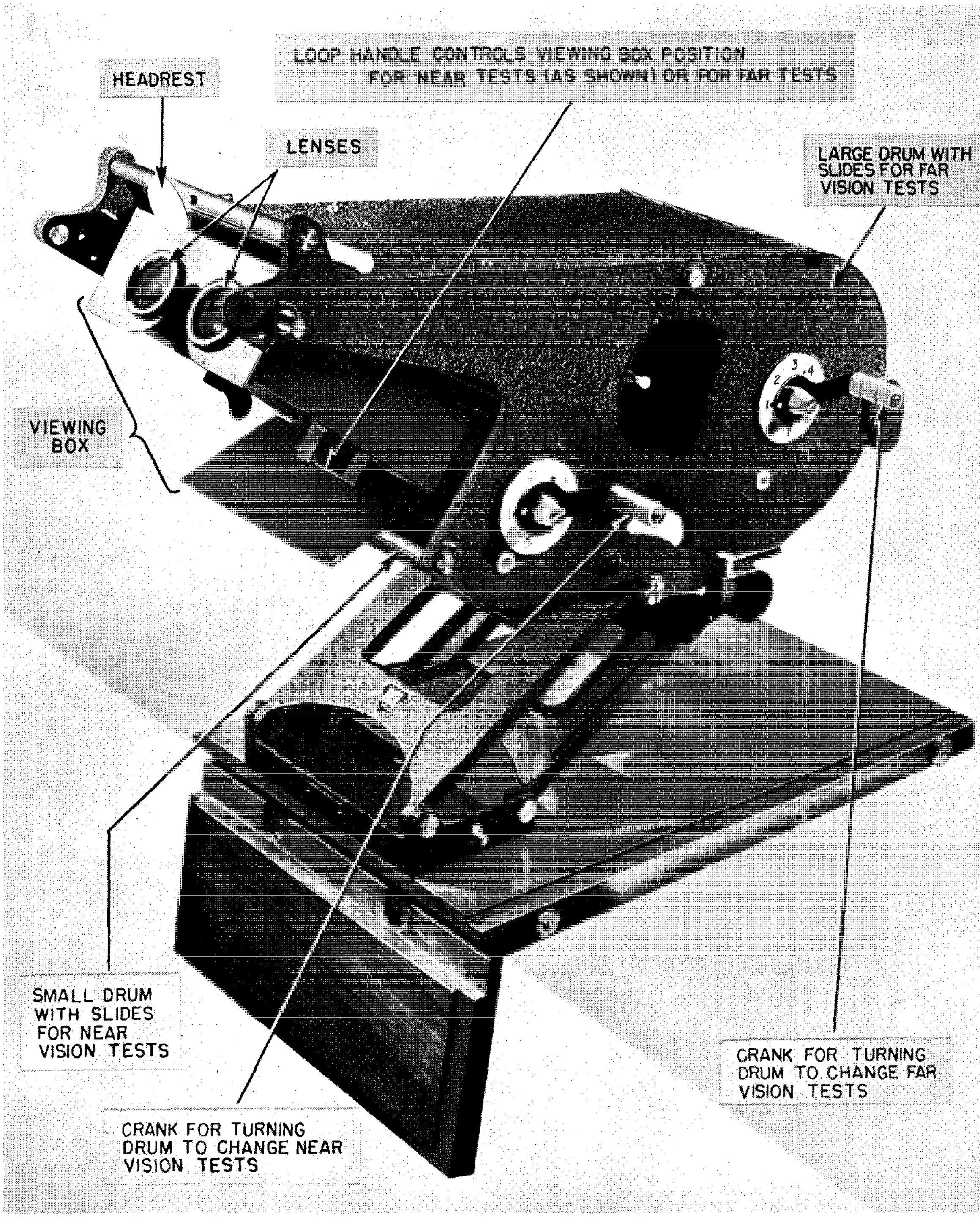


FIGURE 2. Bausch and Lomb Ortho-Rater. View from examiner's position. (Courtesy of Bausch and Lomb.)

Test-retest correlations were lower than might otherwise have been expected (compare columns 1 and 3 in Table 1) because the men in the sample had already been selected to an unknown extent by the Navy on the basis of several visual functions. Therefore, the basic evaluation of the reliability of the Ortho-Rater tests had to be made in terms of the data sum-

[redacted]

marized in the second column of the table—data showing the percentage of test-retest differences which were one score unit or less. As indicated, the Ortho-Rater measures are sufficiently precise and dependable.

The only differences between test means and retest means which were statistically significant were those for right eye and left eye acuity. In both cases, the shift in measures was in the direction of poorer scores on the retest. The standard errors of estimate were small enough, however, to ensure satisfactory prediction in spite of significant test-retest differences between mean scores.

The test of steroscopic acuity given on the Ortho-Rater is a static test. The subject views a series of numbers and indicates which one in a row of five appears nearer to him. There are only six observations of this sort for the subject to make. Reliability of the test, as shown in Table 1, is .64.

TABLE 1. Reliability of the Ortho-Rater industrial vision tests.

	Results on 234 boots tested by a number of Navy examiners: Test-retest correlation	Results on 234 boots tested by a number of Navy examiners: % of men with test-retest difference of 1 or less	Laboratory test of 188 industrial employees by one examiner: Test-retest correlation
Visual acuity, far vision, both eyes	+.49, ±.05	74	....
Visual acuity, far vision, right eye	+.59, ±.04	75	+.89, ±.01
Visual acuity, far vision, left eye	+.59, ±.04	62	+.83, ±.02
Lateral phoria, far vision	+.83, ±.02	84	+.82, ±.02
Vertical phoria, far vision	+.55, ±.05	96	+.62, ±.05
Stereopsis	+.64, ±.04	55	....

8.3.6 Relation between the Massachusetts Vision Test and the Ortho-Rater Tests of Visual Acuity and Phoria

The Massachusetts vision tests and the Ortho-Rater tests of visual acuity and phoria are satisfactory substitutes for each other. They have equivalent reliabilities, and scores on the two tests are closely related as shown in Table 2, which summarizes the results of a special comparison study carried out by experienced test personnel at Fort Eustis, Virginia.[28]

The degree of correspondence between the two tests is shown further by these correlation values.

Visual acuity, right eye	$r = +.80,\ \sigma_r = .02$
Visual acuity, left eye	$r = +.81,\ \sigma_r = .02$
Lateral phoria	$r = +.54,\ \sigma_r = .04$
Vertical phoria	$r = +.44,\ \sigma_r = .05$

It may be noted that the difference between these coefficients of intercorrelation and the previously cited coefficients of reliability for the two sets of tests is an indication of the extent to which the two sets of tests measure different things. The difference is small except in the case of lateral phoria where the difference between the Massachusetts and the Ortho-Rater measurement methods may be of some importance.

TABLE 2. The number and the proportion of subjects passing the Massachusetts and Ortho-Rater tests on visual acuity, lateral phoria, and vertical phoria for distant vision.*

Item		Distant vision: Visual acuity	Distant vision: Lateral phoria	Distant vision: Vertical phoria
N		576	288	288
Pass Mass. Test	*N*	410	273	279
	%	71.3	95.0	97.0
Pass Ortho-Rater	*N*	419	264	268
	%	72.8	91.8	93.0
Pass both	*N*	393	259	266
	%	68.3	90.0	92.4
Agreement, pass or fail	*N*	533	269	273
	%	92.5	92.5	94.6

*Passing scores,			
Mass.:	80% correct	±6 diopters	±½ diopter
Ortho-Rater:	8 or more	2 to 13	4 to 7

[redacted]

8.3.7 The Projection Eikonometer

Among the instruments invented by the Dartmouth Eye Institute to measure stereoscopic vision is one called the projection eikonometer. This instrument, and in particular one test given with it, the stereo-vertical test, was further developed and applied by the Princeton University (Division 7) Field Laboratory at Fort Monroe, Virginia. Its evaluation was continued by the Heightfinder Project N-114 under the Applied Psychology Panel.

In the stereo-vertical test, two lines are projected simultaneously on a vertical screen (see Figure 3). The two images are polarized so that when they are viewed through proper polaroid filters, only one line is visible to each eye. The subject fuses these separately seen images and perceives a single line. During the projection eikonometer test, the two lines are rotated in opposite directions. This causes the perceived line to appear to tilt away from the vertical, either toward the subject or away from the subject. A motor drive returns the lines toward the vertical. The subject presses a button at the instant the stereoscopically fused line appears to be vertical. The angle which the two images on the screen make with each other at the time of the subject's response

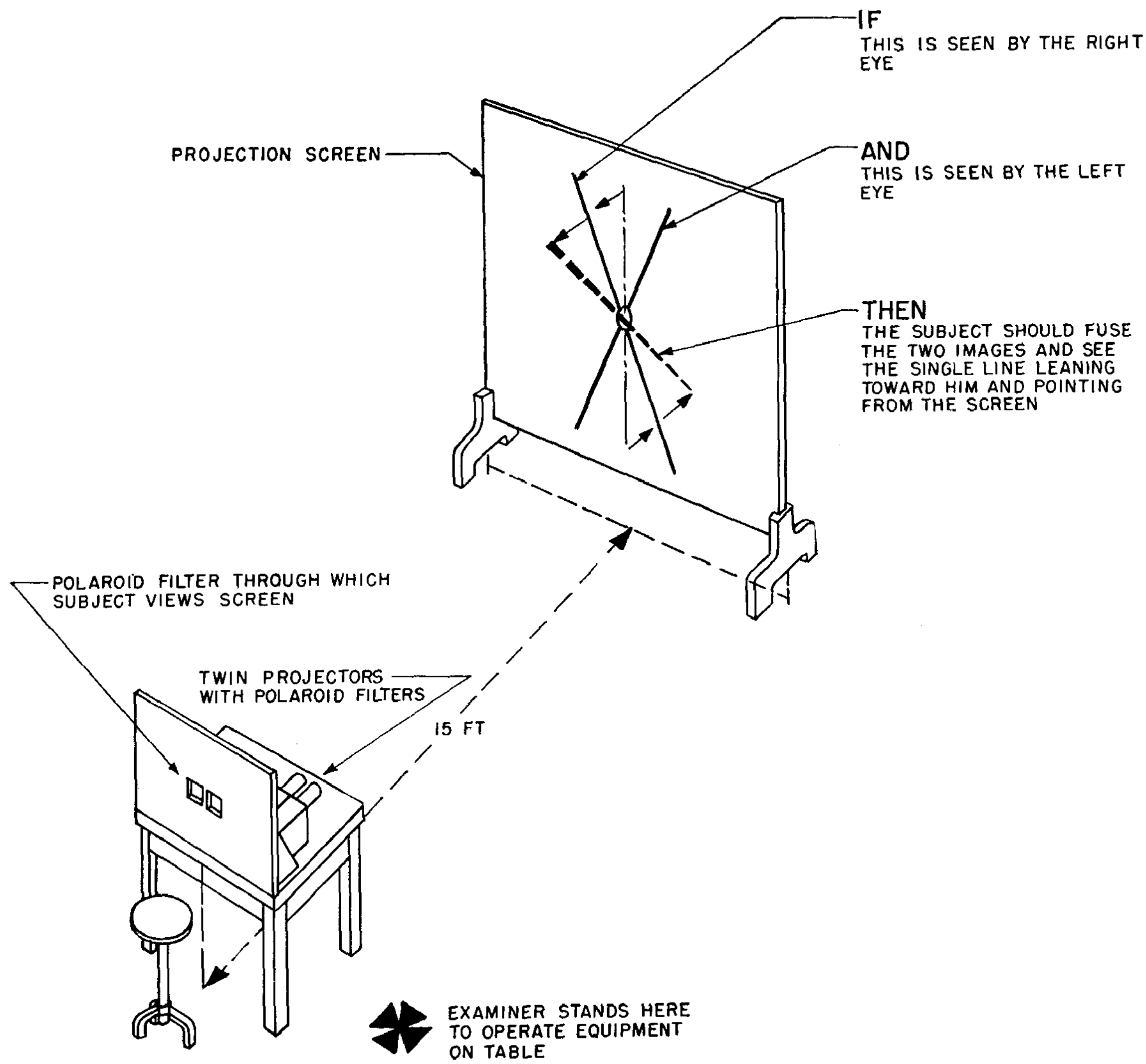


FIGURE 3. Arrangement of the projection eikonometer for testing a single subject.

[redacted]

is recorded. If the subject responds on every trial with zero error (and if he has no cyclophoria), he will always respond when the two images on the screen are parallel and vertical.

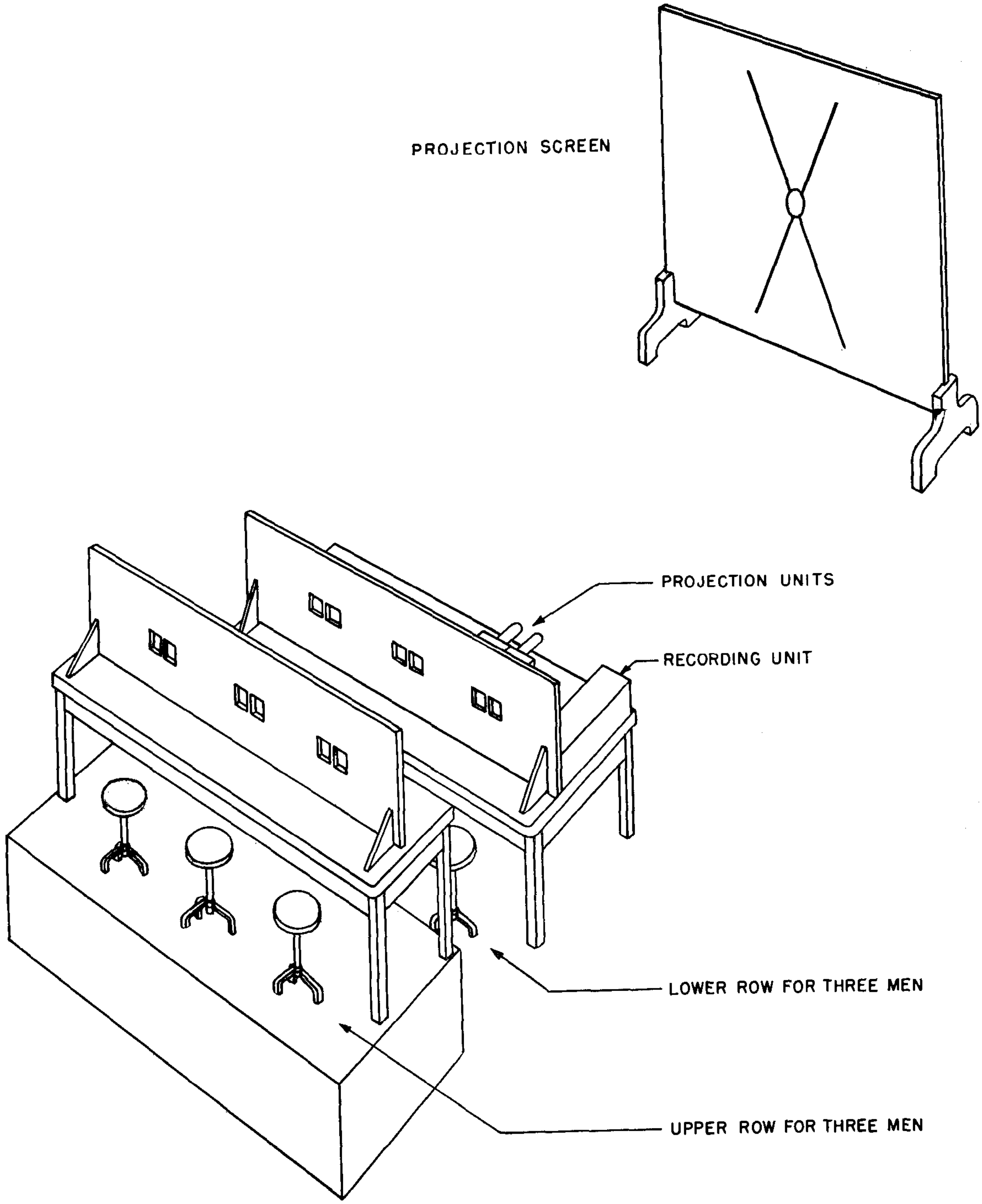


FIGURE 4. Arrangement of the multiple projection eikonometer.

The complete test consists of three "runs" of ten observations each. Successive observations are made as the stereoscopically perceived line alternates in direction of movement. In each "run," then, there are five times when the line is tipped away from the subject initially and moves toward him during the trial, and five when it is tipped toward him initially and

straightens up by moving back. The subject is scored on the basis of the variability of his responses. Raw scores are converted to standard scores by use of a conversion table.

Details of the eikonometer testing and scoring procedures are given in references 6, 8, 19, and 20.

A disadvantage of the original projection eikonometer test was that from 25 to 30 minutes were required to test each subject, and only one subject could be tested at a time. The needed development of the projection eikonometer into a group testing instrument was begun under Division 7 and successfully concluded by Project N-114. The new instrument, known as the multiple projection eikonometer, permits six subjects to view the screen simultaneously. The subjects sit in a small gallery with two rows of three seats each (see Figure 4). While all subjects do not see the screen from exactly the same angle and distance, it was demonstrated that test scores were not affected by these small differences in angle and distance.[11] In the final form of the multiple instrument, each subject has a noiseless key which he presses when the line appears to him to be vertical. The time of the true vertical and the instant at which each key is pressed are recorded on a tape recorder (see Figures 5, 6, and 7 for details of projector and recorder).

With the multiple projection eikonometer, six subjects can be tested in from 40 to 50 minutes. A slightly longer time is spent on the instruction of subjects with the multiple instrument than is required with the single instrument. Otherwise, the administration of the tests is the same.

During the months of January and February 1943, while the multiple projection eikonometer was in operation at Fort Eustis, Virginia, arrangements were made to test a number of men on both the multiple and single instruments.[11] As time in the Testing Center allowed, men were given both tests. One half of the men took the multiple test first (the MS group), and the other half took the single test first (SM group). The experimental data are summarized in Table 3. Means and standard deviations of the distributions of standard scores on the tests are given for the two halves of the population, SM and MS groups, and for both instruments. The standard errors of the means and the critical ratios of the differences between means are also given. All of the critical ratios were small. Scores on the single instrument averaged slightly higher than scores on the multiple instrument, but this difference was not statistically reliable.

TABLE 3. Comparison of scores obtained on the single and multiple projection eikonometers.

Instrument	$N = 99$ SM group		$N = 100$ MS group	
Single eikonometer	Mean = 112.01 σ = 17.8 σ_M = 1.8	C.R. = 0.81	Mean = 113.97 σ = 16.1 σ_M = 1.61	
	C.R. = 0.82		C.R. = 2.05	
Multiple eikonometer	Mean = 110.95 σ = 13.8 σ_M = 1.38	C.R. = 0.68	Mean = 109.62 σ = 14.1 σ_M = 1.41	

The product-moment coefficient of correlation between standard scores on the single and the multiple instrument was +.70 (σ_r = .07). This figure is slightly, but not significantly, less than the test-retest reliability coefficient for the single eikonometer:[12] r = .81 (σ_r = .04). The test-retest reliability of the multiple projection eikonometer[22] is +.78 (σ_r = .03).

From these statistical treatments it is apparent that a test by the multiple eikonometer does not differ from a test by the single eikonometer to a much greater degree than would be obtained with a readministration of either test.

Both the single and the multiple projection eikonometers were routinely used in testing the stereoscopic ability of candidates for training as operators of Army heightfinders or Navy rangefinders.

8.3.8 The Stereoscopic Trainer M2

The stereoscopic trainer M2 is a training device which duplicates the view and task presented to the observer in the heightfinder. It consists chiefly of a stereoscopic presentation

[redacted]

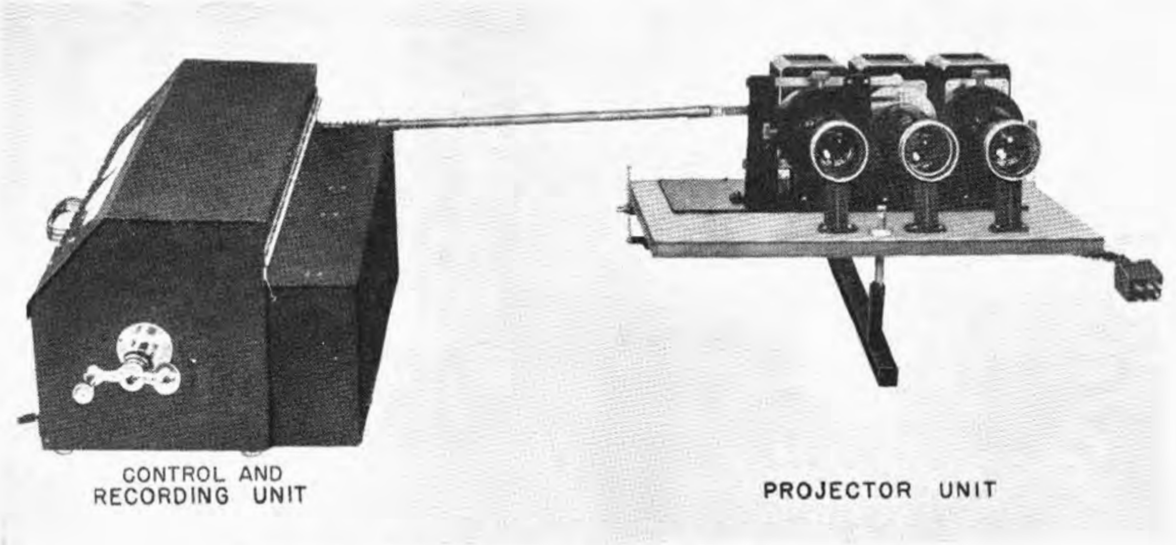


FIGURE 5. The multiple projection eikonometer working units.

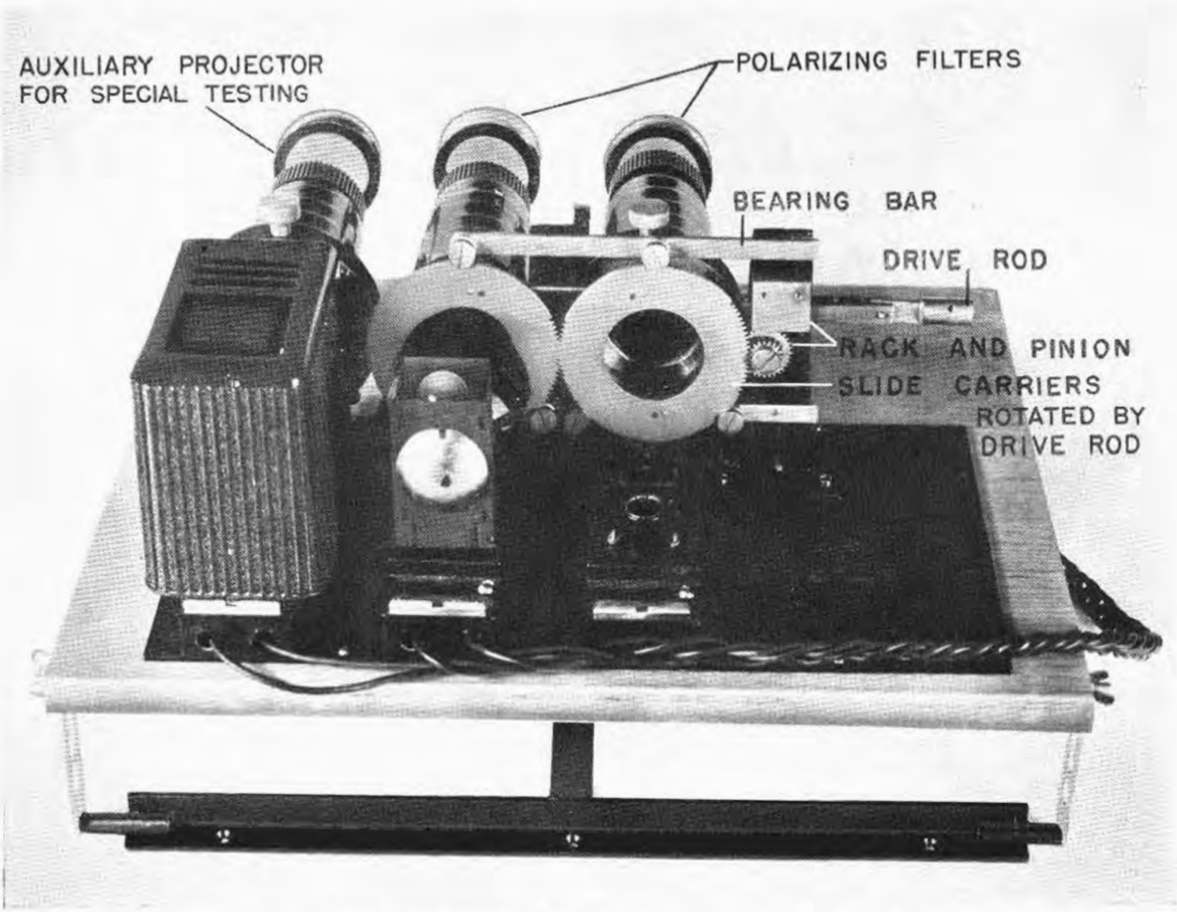


FIGURE 6. Projector unit for multiple projection eikonometer.

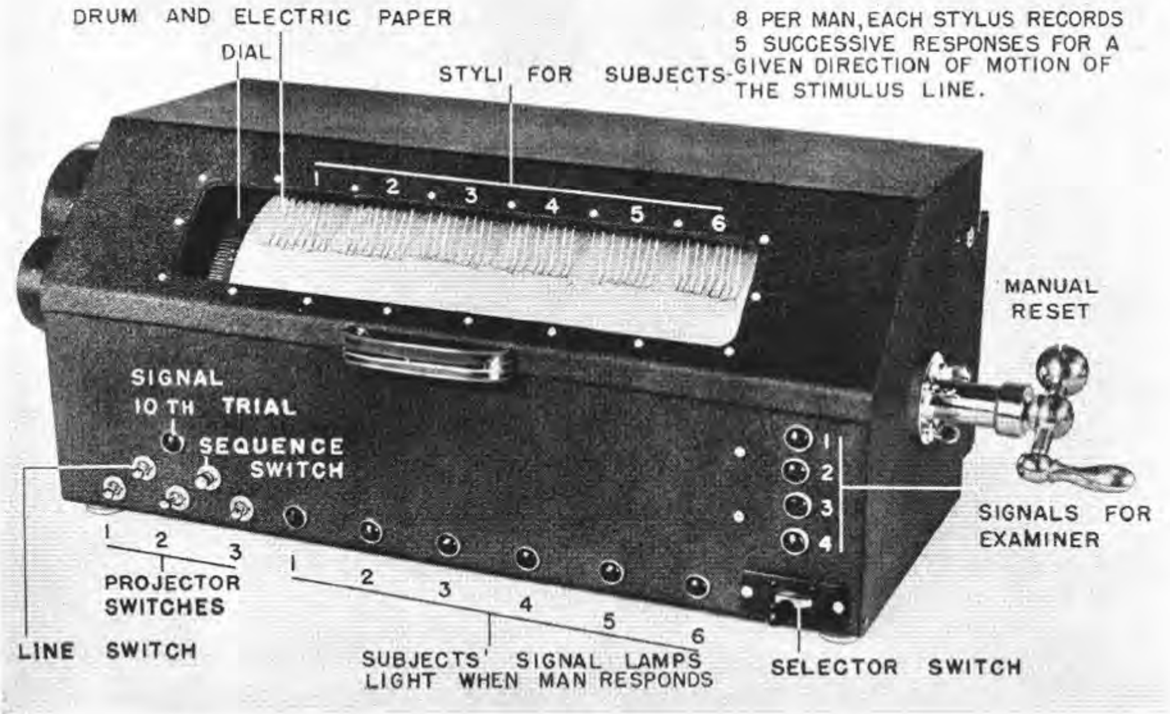


FIGURE 7. Control and recording unit for multiple projection eikonometer.

of an airplane target and a set of reticles. Motion of the range knob on the trainer causes a change in the disparity of the images of the reticles in the two eyes, so that the apparent distance of the target and reticle relative to each other can be varied by definite amounts. The task of the subject is to place the airplane target in stereoscopic contact with the center line of reticle marks.[10]

Six sets of five observations each are made during the test, and the subject is scored on the variability of his judgments. The test-retest correlation for 152 subjects on this test was $+.71$ ($\sigma_r = .05$).[12]

8.3.9 The Vectograph-Pursuit Test

This instrument uses a stationary reticle suspended between two glass plates. On each plate is an airplane target silhouette which has been cut from polaroid material. The axes of polarization of the silhouettes are at right angles to each other. When the targets are viewed through a pair of polaroid spectacles, with axes of polarization of the lenses at right angles to each other, the subject sees a single airplane. It is at the same apparent distance as the reticle if the two polaroid targets are superimposed. But if the glass plates upon which the targets are mounted are moved laterally by motor and cam drive, the subject should see a movement of the target in depth. The subject sits about seven feet from the target-reticle field and turns a "range" knob to offset the movement imparted to the glass plates by the motor and cam. His task is to rotate the knob so as to keep the airplane at the same apparent distance as the reticle. The errors which the subject makes over a two-minute period are summed by an integrator coupled to the cam and gear arrangement.

The test-retest correlation[12] for 46 subjects, unselected for visual acuity or phoria, was $+.75$ ($\sigma_r = .07$).

8.3.10 The Dearborn-Johnston Test

This is a modification of the vectograph-pursuit test which permits the testing of seven

[redacted]

subjects simultaneously. The task of the subject is to count the number of times that the airplane target appears to pass the reticle in depth—i.e., either coming in front from behind or vice versa. The motion of the two polaroid targets to give this fore and aft movement is controlled by a motor-cam arrangement.

The test-retest correlation[12] for 125 subjects, unselected for visual acuity or phoria, was +.67 ($\sigma_r = .05$).

8.3.11 Relations among These Stereoscopic Tests

Reliabilities for the four tests of stereoscopic vision just described have been quoted as varying from .67 for the Dearborn-Johnston to .81 for the projection eikonometer. To be compared with these values are the inter-test correlation data shown in Table 4 and representing test results for an experimental group of 92 men.[12] The intercorrelations between scores on the several tests are conspicuously lower than the reliability coefficients. This suggests that the tests are not measuring exactly the same thing. Each test must be measuring at least some functions which are specific to itself alone. To the extent that a particular test is valid because of the specific factors which it samples, the other tests will probably not be satisfactory substitutes for it in selection.

TABLE 4. Intercorrelations of stereoscopic tests: data† from the Camp Davis validation study.[12]

	PE	M2	VP	DJ
Projection eikonometer (+.81)		+.40*	+.32*	+.21
M2 trainer (+.71)			+.41*	+.16
Vectograph-pursuit (+.75)				+.44*
Dearborn-Johnston (+.67)				

*Significantly greater than zero.
†Population of 92 men, all selected for 20/20 vision and low phoria (reliabilities are given in parentheses).

8.4 THE CAMP DAVIS VALIDATION STUDY[12]

It was pointed out in 8.2 above that the standards which had been set in 1942 for selecting men for admission to the heightfinder school qualified only 4 per cent of selectees. Of 37,500 men screened at Fort Eustis, 27,826 were eliminated by the AGCT and MA2 scores, age, height, and interpupillary distance requirements. The visual requirements of visual acuity and phoria cut the remaining group from 9,674 to 6,242. The two stereoscopic tests (eikonometer and M2 trainer) then passed less than 25 per cent of the men who had qualified in all other respects and cut the final group from 6,242 to 1,474. In view of these facts, the Antiaircraft Artillery Command and Division 7 requested that the Applied Psychology Panel make a further examination of the stereoscopic selection standards to determine whether such high mortality at this stage in the screening was required for successful selection.

8.4.1 Design of the Experiment

The chief deficiencies of the earlier validation experiment were the results of (a) uncontrolled preselection of the subjects and (b) the lack of completely adequate measures of performance on the heightfinder for validation. The present experiment was designed to improve the conditions by which students entered the experimental group and to provide more refined measures of their performance on the heightfinder. Four stereoscopic tests were investigated: the projection eikonometer, the M2 trainer, the vectograph-pursuit test, and the Dearborn-Johnston test.

In order to obtain a representative sample of an untrained Army population, arrangements were made to admit to the Heightfinder Section of the Antiaircraft Artillery School at Camp Davis, North Carolina, a certain proportion of men who did not meet the current standards for selection. The students in Classes 5, 6, and 7 at the Heightfinder Section were selected so that the scatter plot of their paired scores on the projection eikonometer and M2 trainer (the current selection devices) matched as nearly as possible the scatter plot of test scores for a large, random sample of the Army population ($N = 1{,}488$). This procedure guaranteed a suitable spread of ability on the two

instruments and probably an associated spread on the other two stereoscopic tests.

In order to establish criteria of performance on the heightfinder which would be adequate for validating purposes, it was necessary to eliminate accidental and instrumental errors, in so far as possible, and to select measures of performance on the heightfinder which were the best indices of achievement at the end of the three months' course of training at the heightfinder school. The following steps were taken to ensure these ends:

The heightfinders were made more reliable (see Applied Psychology Panel, Volume 2, Chapter 24) by:

1. Indoctrinating students in maintaining the proper charge of helium.
2. Training students to calibrate the instruments in a more satisfactory way.
3. Providing end-window stops to reduce parallax errors.
4. Providing instrument sunshades to reduce temperature stratification.
5. Determining an accurate interpupillary distance measure for each student and training him in making interocular settings using templates.

The accuracy of achievement data was assured by:

1. Putting the instruments into proper adjustment.
2. Obtaining accurate measurements of the altitudes of these missions from phototheodolite records which were synchronized with the heightfinder observations.
3. Devising and providing simplified record forms for the students, from which both accuracy and variability scores could be obtained.

8.4.2 The Experimental Population

The validation population was drawn from Classes 5, 6, and 7 at the heightfinder school. It included a total of 92 men, each of whom satisfied all the following criteria:

1. He must have been tested on the four stereoscopic tests.
2. He must have had no previous training on the heightfinder.
3. His training and final examination must have been conducted on a heightfinder which was in excellent condition and adjustment.
4. His records must have contained adequate criterion measures of his heightfinder performance.

To the scores of these 92 men on each of the stereoscopic tests, the chi-square test of independence was applied in order to determine whether the score distributions for the experimental population were comparable to those for larger, randomly selected Army populations. The p values were as follows: eikonometer, 0.37; M2 trainer, 0.92; vectograph, 0.02; and Dearborn-Johnston, 0.20. From the high p values for at least three of these tests it follows that the distributions of test scores for the experimental sample did not differ significantly from the distribution found in the large sample drawn at random from the Army population. For the vectograph test with a p value of 0.02, there is some doubt that the distribution of scores in the experimental sample was comparable to the distribution within the random sample.

8.4.3 Performance Criteria on the Heightfinder

Reliable and accurate measures of the proficiency with which observers can use the heightfinder are essential to an experimental program for developing tests to select satisfactory heightfinder operators. Such measures of performance must distinguish among those men who are good, bad, and indifferent operators of the instrument. Only with such satisfactory measures of performance on the heightfinder is it possible to determine whether the tests administered before training predict successes and failures.

The criterion scores for the Camp Davis experimental population were based upon the last 22 aerial courses of the final examination. "True" target positions were obtained from phototheodolite records. For each man, the 22 scores were all obtained within a 10- to 14-day period. Each man kept records of the range and height readings which he obtained with

his instrument. From these data and from the altitude measurements obtained by the Records Section from its phototheodolite readings, the observers computed measures of the accuracy and variability of their readings.

The following five measures of performance on the heightfinder were obtained from the final examination data:

1. Variability score in UOE.[e]

   The variability score is a measure of the variation in the readings taken by the heightfinder operator. For the 22 courses, this score is the median of the variability scores for the individual courses.

2. Course error score in UOE.

   This score is an accuracy score. On a single course, the course error is the median algebraic error, and is expressed in UOE. The course error for the 22 courses is the median (disregarding algebraic sign) of the errors for the individual courses.

3. The "hit" score in per cent.

   This scoring method divides each course into 5 parts. Those parts of an individual course where the median reading falls within the limits of 1 UOE behind true target position and 2 UOE in front of true target position are considered to be "hits." The criterion "hit" score is the total number of "hits" obtained for the last 22 courses of the examination.

4. Percentage of "good courses" score.

   Courses which show a variability of 4 UOE or less and a course error of 2 UOE or less are considered "good courses." "Good courses" are treated as a percentage of the total number of courses. This score emphasizes the nature of data which are good in terms of gun director function.

5. The sum of the course error and the variability scores in UOE.

   In terms of the relative contributions of bias and variability to director performance, it turns out that a sum of course error and variability scores (as measured by methods used in the experiment) is a good index of the relative inadequacy of the data for a gun director.

Test-retest reliability coefficients for these five criterion scores were obtained by correlating scores computed for the first 11 courses with scores computed for the last 11 courses of the final examination. The uncorrected coefficients are presented in Table 5. The magnitude of these correlations indicates satisfactory reliabilities for the performance criteria.

TABLE 5. Reliability coefficients for the criterion scores.

Criterion	Reliability
Variability	0.82
Course error	0.68
Sum of variability and course error	0.87
"Hits"	0.58
Per cent of good courses	0.67

Three of the criterion scores, the variability score, the course error score, and the sum of the two, are "error" scores. "Error" scores lend themselves more easily to exact statistical treatment after they are transformed to logarithms, because the logarithms of these scores are more normally distributed than are raw scores. For this reason, logarithms were used whenever correlations involving one of these variables were computed. The other two scores, the "hit" score and the per cent of good courses, were expressed as percentages, and the raw percentages were used as the criterion scores in correlation analysis.

TABLE 6. Intercorrelations of the criterion scores.

	Course error	"Hits"	Per cent of good courses	Sum of variability and course error
Variability	0.80	−0.80	−0.73	0.97
Course error		−0.25	−0.69	0.92
"Hits"			0.84	−0.85
Per cent of good courses				−0.75

Table 6 gives the intercorrelations among the five criterion scores. The magnitudes of these correlations indicate that each of the

[e] UOE is the usual abbreviation for "unit of error." One UOE is a disparity difference of 12 seconds of arc at the eye of the observer.

[redacted]

criterion measures was closely related to every other criterion measure.

In the preliminary statistical analysis of the relations of stereoscopic test scores to criterion scores, all five of the criterion scores were used at a number of different cutoffs. Because of the relation of all the criteria to each other, it soon became apparent that it would be satisfactory to consider in detail the results for only one or two criteria. One is chosen for treatment here: the sum of the variability and course error with a cutoff at 9 UOE. This score is highly related ($r = +.97$) to the variability score which was the conventional measure of performance in use for graduation at the heightfinder school. It is the score which theoretically defines a man's performance in terms related to the probability of his success as an integral part of a successful antiaircraft unit. It has a component which is an accuracy score, and is very highly correlated with the "hits" score which is based on absolute accuracy of the observer's readings.

8.4.4 Correlation of Stereoscopic Test Scores with the Criterion

Table 7 shows the product-moment correlation coefficients between stereoscopic test scores and the performance criterion. The coefficients are negative because the test scores increase with goodness, while the criterion scores increase with performance errors.

TABLE 7. Product-moment correlations of test scores with the performance criterion.

	Sum of variability and course error
Projection eikonometer	−0.50
M2 trainer	−0.43
Vectograph-pursuit	−0.39
Dearborn-Johnston	−0.29

Although none of these correlations was high, the tests may be successful in weeding out the worst men. So further analysis of the efficiency of selection was made using chi-square tests and "cost" comparisons. By way of illustration, assume that the projection eikonometer test is proposed for use in selection and that a passing score of 105 is being considered for adoption. The questions to be answered in deciding on the usefulness of the test are these: (1) Does the test distinguish reliably between good men and poor men when 105 is used as the passing score; i.e., are the men who pass a significantly better group by a chi-square test than those who fail? (2) How many men are eliminated by using 105 as a passing score; i.e., what is the "selection cost"? (3) How many men get to school but fail the course; i.e., what is the "instructional cost"?

For the case suggested, pertinent data are given in Table 8 and Figure 8.

TABLE 8. Chi-square table for eikonometer passing score at 105. ($N = 92$).

	Men with eikonometer score 104 or below	Men with eikonometer score 105 or above
Men who passed criterion	15	46
Men who failed criterion	17	14

Chi-square analysis shows that the data in Table 8 do depart significantly from a chance distribution. The test does distinguish reliably between good men and poor men. Of the men eliminated by the test, more would fail the course than pass it (see left half of the table). Of the men passed by the test, three-fourths would pass the course (see right half of the table).

Diagrammatically, instructional and selection costs can be shown in figures like those in Figure 8.

Note that if 100 trained men are required, one would have to start with 200; 70 would fail the selection test, 130 would take training, and 30 would fail the course. But if no selection had been used 156 men would have to take training; 56 men would fail, leaving 100 graduates. Thus, selection saves the instruction time for 26 failing men.

The best cutoffs for the individual tests and combinations of tests against the sum of the variability and course error scores are summarized for comparison in Figure 9. These

tests and combinations of tests at the indicated cutoffs give the best prediction of successful operation of the heightfinder together with the least waste in terms of selection cost. All these methods of selection were statistically reliable in distinguishing between good and poor heightfinder operators.

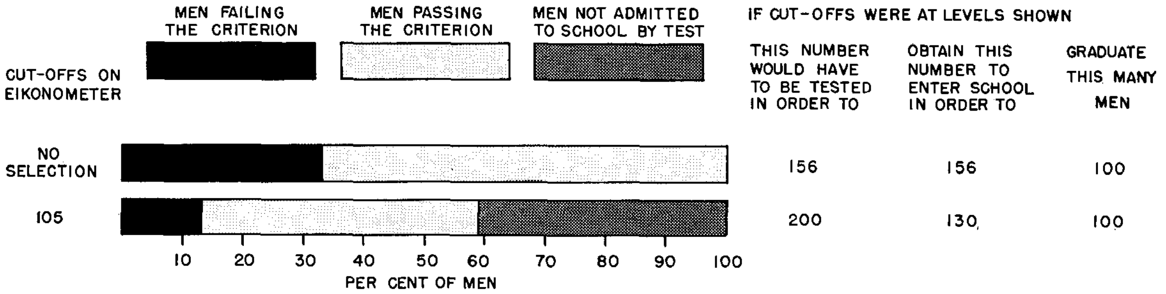


FIGURE 8. Selection data for eikonometer passing score at 105.

Of first importance to the efficiency of the heightfinder school is the number of men who fail to meet the graduation requirements after being given instruction at the school. This might be called *instruction cost*. Instruction cost is high when the selection of school candidates is inadequate. Instruction cost is indicated in the second column at the right in Figure 9, which shows how many men would have to be admitted to the school in order to graduate a class of 100.

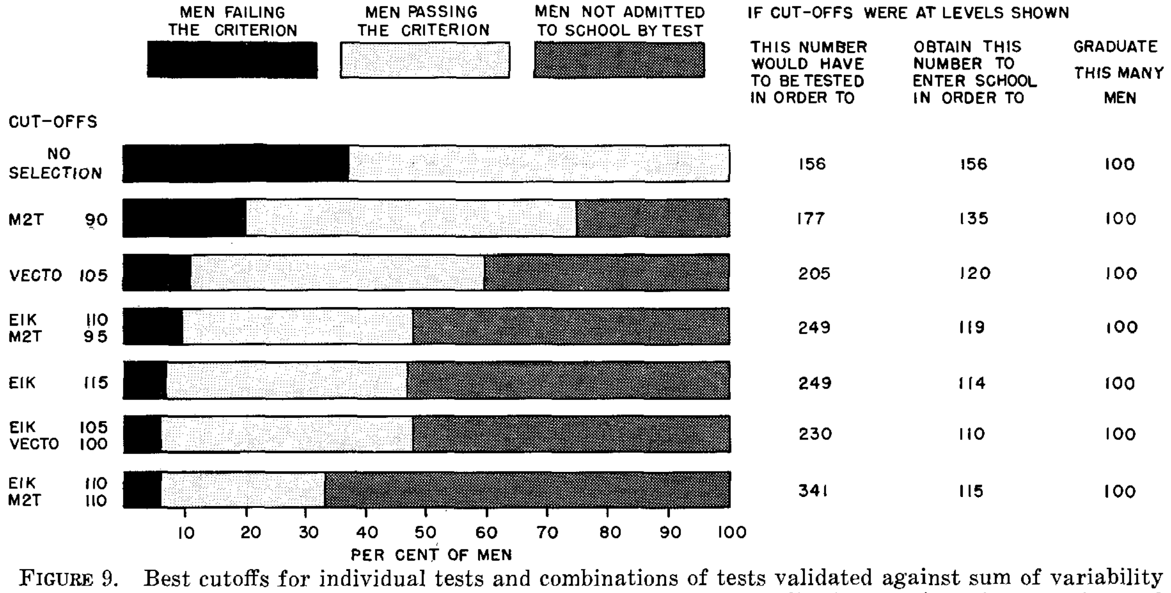


FIGURE 9. Best cutoffs for individual tests and combinations of tests validated against sum of variability and course error scores. Based on criterion population of 92 men, all with passing visual acuity and phorias.[12]

The value of a selection program may be judged by how well it reduces instruction cost with a reasonable amount of selection cost. The *selection cost* refers to the number of men who must be tested in order to obtain a certain number of "good" men for training at the heightfinder school. This figure is given in column 1 to the right of the figure. Thus, 249 men have to be tested to pass 114 men on the eikonometer with a score of 115.

Thus in deciding which of these tests, in which combinations, and at what cutoffs, should be recommended as standard Army practice in selecting men for training as heightfinder

operators, most attention should be given to the instruction cost and the selection cost. Note in Figure 9 that the eikonometer at 115 is equivalent in instruction cost to the eikonometer and M2 trainer tests both at 110 but is far superior in having a low selection cost. But one should also consider amount of time required to administer the test or tests per man and the availability of instruments for the administration of the tests.

All the tests considered in Figure 9 were individual tests, although the eikonometer was shortly afterwards converted to a multiple testing instrument which tested six men at once. The eikonometer requires about 25 minutes per subject as a single instrument. The M2 trainer test can be administered in about 27 minutes to an individual subject. The vectograph-pursuit test takes about 8 minutes per subject.

At the time of this study there was only one vectograph-pursuit instrument in existence. M2 trainers were available to any group in the Army that might be faced with the problem of selecting candidates for training as stereoscopic heightfinder operators. The Stereoscopic Testing Centers and the mobile units of these centers were adequately equipped with projection eikonometers.

On the basis of the experimental results and the considerations of testing time and equipment availability it was concluded that:

1. All combinations of tests should be eliminated from consideration, since single tests did about as well as combinations and required less time to administer.

2. The eikonometer was the most generally satisfactory of any of the tests, but the M2 trainer might be used if an eikonometer were not available.

3. The vectograph-pursuit apparatus showed real promise, particularly in its low selection cost. If it became necessary to expand the testing program, it should be given consideration.

8.4.5 Recommendations

In the light of these conclusions, it was recommended to the Antiaircraft Command that (1) the test on the stereoscopic trainer M2 be dropped from the selection battery (as given in 8.2.1 above) and that the passing score on the projection eikonometer be raised from 110 to 115, and (2) for field or other use where the projection eikonometer was not available, men be accepted for training who score the highest on the M2 trainer test, none being accepted who score below 90.[12]

These recommendations were accepted, and the changed requirements were immediately put into effect by the Antiaircraft Command.

8.4.6 Follow-up Study of the Validity of the Projection Eikonometer

After the new requirement of a score of 115 on the projection eikonometer was put into effect, a follow-up study of the effect of this standard in weeding out the potentially inapt heightfinder operators was made.[21] Students of eight consecutive classes of stereoscopic heightfinder operators at Camp Davis, North Carolina, were studied to determine the frequency with which men scoring above 115 on the projection eikonometer failed the heightfinder course. Results for this new group of men were practically identical with the results of the earlier study. The percentage of men failing on the variability score (score over 6 UOE) had been 7.1 per cent in the original validation group. It was now 7.9 per cent. The percentage who failed on the combined variability and course error score was 11.9 per cent in the validation group and in the new classes was 10.3 per cent. The critical ratios of both differences were small.

The value of the projection eikonometer as a selection device is shown by comparing the percentage of men who failed the course with the much larger percentage who would have failed it if the projection eikonometer were not used as a part of the selection procedure. With no selection on the projection eikonometer, about 33 per cent would have failed on the variability score and about 36 per cent on the combined score. The use of the projection eikonometer reduced these failure ratios to 7.9 per cent and 10.3 per cent respectively.

[redacted]

8.5 THE FORT LAUDERDALE VALIDATION PROGRAM[22]

Applied Psychology Panel Project N-114 was set up at the request of the Navy at NTS, Fort Lauderdale, Florida, to assist the school in the selection and training of its fire control students. In the course of its work, N-114 administered stereoscopic and other tests to all incoming students. Finally, these test scores were correlated with student achievement. The report on these correlations[22] concludes with recommendations for the selection of future rangefinder operators.

8.5.1 Plan of the Experiment

The experiment was carried out under conditions closely resembling those which prevailed during the early validation work by Division 7 at Fort Monroe. The men in school classes were not specially selected, but it is not known what selection factors may have operated in determining their assignment to the school. Normal school practice determined the degree of maintenance of the school rangefinders and the precision of the radar reference ranges. Criterion scores were based upon variability scores (see Section 8.4.3) obtained during the school's final examination for each class.

The experiment included the examination of a series of tests which might be used in predicting school grades on theory work. These tests were some standard Navy pencil-and-paper tests (GCT, MAT, MKM, MKE, Reading, Arithmetic), some tests developed by Applied Psychology Panel Project SC-70 for possible use in selecting radar operators, and an interest test.[16]

The method of analyzing the results was the same as that used in the Camp Davis experiment: chi-square tests of cutoff reliability and considerations of selection cost and instruction costs.

8.5.2 Prediction of Rangefinder Operation from Visual Test Scores

As in the Division 7 studies, very few cases of men with poor visual acuity or excessive phoria were encountered in the school populations. The retention of selection standards for acuity and phoria therefore remained a matter of convenience for other testing.

TABLE 9. Application of visual test scores in predicting rangefinder variability scores.[22]

	Number of men who must be tested to get	This number of men to enter school in order to	Graduate this number of men
No selection	136	136	100
Visual acuity—far vision score of 9 or more on O-R (20/22 vision)	145	130	100
Vertical phoria—scores 4 to 7 on O-R	136	134	100
Lateral phoria—scores 4 to 12 on O-R	140	131	100
Visual acuity and phorias as above	150	127	100
Projection eikonometer score of 50	193	119	100
Projection eikonometer score of 58	336	114	100
Ortho-Rater depth test score of 4	196	125	100
Ortho-Rater depth test score of 5	231	118	100
VA+phoria+PE at 50	207	116	100
VA+phoria+PE at 58	362	110	100
VA+phoria+O-R depth at 4	204	120	100
VA+phoria+O-R depth at 5	243	116	100

The selection cost and instructional cost of the different tests used are given in Table 9.

This table shows that the most efficient selection from the standpoint of instructional cost is the set of visual tests recommended on the basis of the Camp Davis work. The selection cost of the eikonometer at 58 (equal to 115 in Army scores) was, however, somewhat higher than

[redacted]

that reported for the Davis study. If greater instruction cost can be tolerated, substitute stereoscopic acuity standards might be used: the projection eikonometer score at 50 or the Ortho-Rater depth test at 5. The eikonometer score would be preferred because of lower selection cost, in terms of men, but the Ortho-Rater test has the advantage of low testing time, and hence low selection cost in terms of time.

8.5.3 Prediction of Classroom Grades from Pencil-and-paper Tests

The classroom grade—called the General Principles Grade—used at NTS, Fort Lauderdale, was a composite of grades in Rangefinder Principles, Fire Control and Gunnery, Fire Control Problems, and Radar Principles. The relationship of this composite grade to pencil-and-paper tests was examined.

Good selection was possible with the General Classification Test and with the Arithmetic Test (see Table 10). Mechanical Aptitude Test scores, Reading scores, and Electrical Knowledge Test (MKE) scores were related to the criterion, but good selection was possible only with a high selection cost. The Mechanical Knowledge Test (MKM), Spatial Relations Test, Oscilloscope Operator's Test, and Activities Preference Test showed by chi-square tests no significant relations to the criterion.

TABLE 10. Application of pencil-and-paper test scores in predicting classroom grades for rangefinder operators.[22]

	Number of men who must be tested to get	This number of men to enter school in order to	Graduate this number of men
No selection	122	122	100
GCT passing score at 45	135	118	100
GCT passing score at 50	160	112	100
GCT passing score at 55	225	104	100
Arith. Test passing at 45	146	109	100
Arith. Test passing at 50	181	106	100
Arith. Test passing at 55	292	103	100
GCT at 50 and Arith. at 50	193	105	100
GCT at 50 and Arith. at 45	166	109	100

8.5.4 Predicting Overall Student Performance

In the prediction of combined classroom and rangefinder performance, two batteries of tests proved to be about equally effective. Those with the best cutting scores for each were the following:

I GCT 50, Arithmetic 45, Visual Acuity 9, Vertical Phoria 4-7, Lateral Phoria 4-12, Eikonometer 58. To graduate 100 men, 521 men must be tested. One hundred and seven survive the above tests and enter the school. Seven can be expected to fail.

II GCT 45, Arithmetic 55, Visual Acuity 9, Vertical Phoria 4-7, Lateral Phoria 4-12, Ortho-Rater Depth Score 5. To graduate 100 men, 584 men must be tested. One hundred and eight survive the tests and enter the school. Eight can be expected to fail.

These two batteries of tests were recommended to the Navy for use in qualifying men for rangefinder operation training.

8.6 OTHER DATA BEARING ON THE SELECTION OF STEREOSCOPIC OPERATORS

In laboratory research work carried out under Division 7 contracts, it was demonstrated that women are as good as men at stereoscopic

operation.[7] It was also shown that chromatic dispersion in the eye could be serious enough in some individuals to disqualify them from rangefinder operation.[5]

8.7 RELATION BETWEEN VISUAL ACUITY AND STEREOSCOPIC ACUITY

The relation between visual acuity and stereoscopic acuity was examined in order to determine whether a lowering of the visual acuity standards would result in an increase in the number of men passing the stereoscopic tests, thus increasing the pool of men eligible for training as stereoscopic heightfinder observers.[13]

One thousand and fifty-two soldiers were permitted to take the tests of stereoscopic vision at the Stereoscopic Testing Center at Fort Eustis, Virginia, even though they failed to meet the usual standard of visual acuity of at least 20/20 in each eye (unaided). Analysis of the results showed that although the elimination of the test for visual acuity would produce a slight increase in the number of men who qualified on the stereoscopic tests, the burden of testing would be greatly increased. A large majority of men with vision poorer than 20/20 failed the stereoscopic tests. Thus, relaxing the visual acuity requirement would not be economical in testing time.

When a visual acuity standard of 20/20 in each eye (unaided) was applied, 27.6 per cent of the men secured a standard score of 110 or higher on the projection eikonometer. If the visual standard were dropped to include all those who make no more than one error on the 20/30 chart, the percentage making a standard score of 110 or higher on the projection eikonometer would increase only 5.6 per cent, to a total of 33.2 per cent. To obtain this small increase in the number of men with satisfactory stereoscopic ability, it would be necessary to examine 82 per cent, instead of 64 per cent, of the total Army population which qualifies on the basis of age, height, AGCT, and MA tests.

In view of these findings, it was recommended that the visual acuity standards (Section 8.2.1) for selecting men for the heightfinder school should not be relaxed.

Because it is of general psychological interest, it should be added that the stereoscopic test scores of those men who missed more than 16 of the 20/20 letters on the visual acuity test (25 per cent of the test population) were significantly inferior to the test scores of the men who missed no more than 5 of the 20/20 letters (64 per cent of the population).

8.8 CORRELATIONS AMONG TESTS OF STEREOSCOPIC VISION

There have been three sets of experiments which have contributed data[4, 12, 24, 25, 26, 28] on the intercorrelation between tests of stereoscopic vision. For purposes of direct comparison, the results of these studies are presented in Tables 4, 11, 12, and 13. The conditions of each of the experiments are summarized in the table footnotes.

TABLE 11. Intercorrelations of stereoscopic tests: data† from Tufts College studies, Division 7.[24, 25, 26]

	Tufts trainer	Vectograph-pursuit	Botts test	Keystone test	Wulfeck test
M2 trainer (+.71)	.47*	.50*	.22*	....	.35
Tufts trainer (+.55)		.54*	.34*	....	−.11
Botts test (+.79)		.32		+.69*	−.39

*Significantly greater than zero.
†Population of 56 men (all 20/20 vision) and 32 women (no glasses). (Population smaller on Wulfeck test and vectograph test.) Reliabilities are given in parentheses. Tufts trainer, M2, and vectograph-pursuit were dynamic tests; Botts, Keystone, and Wulfeck tests were static tests.

For a more complete description of those tests which are new to the reader, reference to the original reports is recommended. The tests

[redacted]

involved are of two types—"dynamic" tests, where the subject sees a target move or moves it himself, and "static" tests where the subject merely makes a judgment about the relative stereoscopic position of two or more figures.

TABLE 12. Intercorrelations of stereoscopic tests†: Fort Eustis data for 192 men with vision of 20/20 or better.[28]

	M2	VP	DJ	W	O-R	Keystone
Projection eikonometer	.42*	.23*	.38*	−.29*	−.39*	−.34*
M2 trainer (M2)		.36*	.41*	−.49*	−.38*	−.41*
Vectograph-pursuit (VP)			.31*	−.14*	−.31*	−.30*
Dearborn-Johnston (DJ)				−.31*	−.29*	−.28*
Wulfeck (W)					.29*	.33*
Ortho-Rater (O-R)						.47*

*Significantly greater than zero.
†Projection eikonometer, M2 trainer, vectograph-pursuit, and Dearborn-Johnston tests were all dynamic. Wulfeck, Ortho-Rater, and Keystone tests were all static. Negative coefficients in the dynamic vs static block of Tables 12 and 13 show positive relations because the dynamic tests were graded as "error" scores and the static were graded as "right" scores.

The correlations which are significantly greater than zero (confidence level of 5 per cent) are marked with an asterisk in the tables.

The important fact about the intercorrelations shown in Tables 4, 11, 12, and 13 is that, although many of them are significantly greater than zero, none is high enough to permit reliable prediction from one test to another. From this and the argument already presented in Section 8.3.11, it is safe to conclude that the data do not indicate that any one of the tests is a substitute for another.

TABLE 13. Intercorrelations of stereoscopic tests†: Fort Eustis data for 96 men with less than 20/20 vision.[28]

	M2	VP	D-J	W	O-R	Keystone
Projection eikonometer	.37*	.46*	.34*	−.09	−.11	−.14
M2 trainer (M2)		.53*	.21*	−.21*	−.27*	−.42*
Vectograph-pursuit (VP)			.36*	−.19	−.21*	−.31*
Dearborn-Johnston (DJ)				−.51*	−.06	−.12
Wulfeck (W)					.10	.39*
Ortho-Rater (O-R)						.29*

*Significantly greater than zero.
†Projection eikonometer, M2 trainer, vectograph-pursuit, and Dearborn-Johnston tests were all dynamic. Wulfeck, Ortho-Rater, and Keystone were all static. Negative coefficients in the dynamic vs static block of Tables 12 and 13 show positive relations because the dynamic tests were graded as "error" scores and the static were graded as "right" scores.

Thus, it is apparent that validation results obtained for one test in a selection experiment cannot be used to indicate what might happen in a similar experiment with another test.

8.9 FUTURE WORK ON VISUAL TESTS

Neither the medical interest nor the scientific interest in problems of vision has yet led to the production of a completely satisfactory set of visual tests. Some visual functions can be measured fairly reliably, but for other functions (see also Chapter 9) the best available instruments provide scores of very low reliability. Tests which purport to measure the same functions show low correlations with each other.

A fairly extensive program of research on visual requirements and visual tests would have a very good chance of producing results of great value to the Army and the Navy. The research should include:

1. A thorough experimental and statistical analysis of the various visual functions and abilities and of the relations between them.

2. A detailed study of the requirements of a number of representative military tasks in terms of each type of visual ability.

[redacted]

3. The construction of new tests and the improvement of existing tests to measure each type of visual ability. Several tests for each type should be available. By the regular procedures of test standardization and validation, the best tests could then be selected and validated.

Because many of the visual tests in use today were devised from a medical point of view and were intended to measure the degree of visual impairment or defect, they have not always proved too satisfactory as predictors of success in jobs which involve visual ability. The tests and testing procedures described in this chapter were developed especially for use in selecting good men. Approaching the whole visual testing problem from the standpoint of predicting visual proficiency may lead generally to tests of greater usefulness in military selection. The effort is well worth making.

[redacted]

Chapter 9

SELECTING NIGHT LOOKOUTS

By *William E. Kappauf, Jr.*[a]

Summary

THE APPLIED PSYCHOLOGY PANEL and the National Research Council's Committee on Human Aspects of Observational Procedures cooperated with other groups in the development of two adaptometers, the NDRC adaptometer, Model III, and the modified Rostenberg adaptometer.

Two studies on the prediction of night lookout performance both showed that the tests of night vision being used had low reliability and very low validity.

9.1 THE PROBLEM OF SELECTING LOOKOUTS

For 99.9 per cent of the time the job of night lookout is a thankless one. It is tedious, uninspiring, poorly rewarded work. Men assigned to the job are often those left over when all other billets have been filled. They cannot be rated as long as they stay at the job.

These conditions have meant that, during the war, night lookouts were generally chosen by negative selection rather than by positive selection. Only one particular capacity has been sought in the lookout—good night vision, or at least evidence of no night blindness. This in itself might be an adequate selection criterion were it not for the fact that most night vision testing devices, usually called adaptometers, are of very low reliability. Men taking the test usually improve their scores when retested and do not rank in the same order of ability on the second test.

It is therefore not surprising that a great amount of research time during the war should have been devoted to the development of a more satisfactory and dependable adaptometer and to answering the fundamental question as to whether an adaptometer is a good selection device for night lookouts. The Applied Psychology Panel participated in this work. It assisted in the design of two adaptometers and conducted field experiments in which tests of night vision were evaluated for their usefulness in predicting night lookout performance.

9.2 THE NDRC ADAPTOMETER

The National Research Council's Committee on Human Aspects of Observational Procedures in cooperation with other OSRD sections and the Navy developed the NDRC adaptometer, Model III.[1] This adaptometer is similar to the radium plaque adaptometer officially adopted for use by the U. S. Navy, except in its source of illumination. An incandescent bulb rather than a radium plaque provides the necessary illumination. The man being tested fixates a small red cross and reports the orientation of a peripherally located T-shaped figure which is flashed on for brief intervals. The observation distance is five feet. Testing can be carried out at any of five different brightness levels (see Figures 1 and 2).

This instrument was not adopted for routine use by any Service.

9.3 THE MODIFIED ROSTENBERG ADAPTOMETER

In response to a request from the Technical Section, Classification and Replacement Branch, Office of the Adjutant General, U. S. Army, a representative of the Applied Psy-

[a] This chapter is based on the work of NDRC Project N-115 and work done by the National Research Council's Committee on Service Personnel—Selection and Training.

The Applied Psychology Panel was only one of several groups seeking better methods of selecting men for duty at night. Others working on this problem included the Adjutant General's Office, the Army Air Forces, the Armored Medical Research Laboratory, and the Medical Research Laboratory, U. S. Submarine Base, New London, Connecticut. Some of the reports from these other agencies claimed more positive results than those described in this chapter.

chology Panel Project N-115 went to Camp Blanding, Florida, to assist in testing an adaptometer designed by Major Rostenberg. Only preliminary tests were run, but as a result of this experience in night vision testing, the personnel of Project N-115 later constructed two modified and improved models of the Rostenberg adaptometer.[5] One of these modified instruments was transferred to the Naval Medical Research Laboratory, Bethesda, Maryland, for test and further study. No report on the results of this test has been received.

The modified Rostenberg adaptometer has a radium plaque as its source of illumination. Test objects of any shape or size up to 16 inches in diameter can be shown in any of 8 positions. The adaptometer is used at a distance of 20 feet. Thus, the subject uses distance vision as

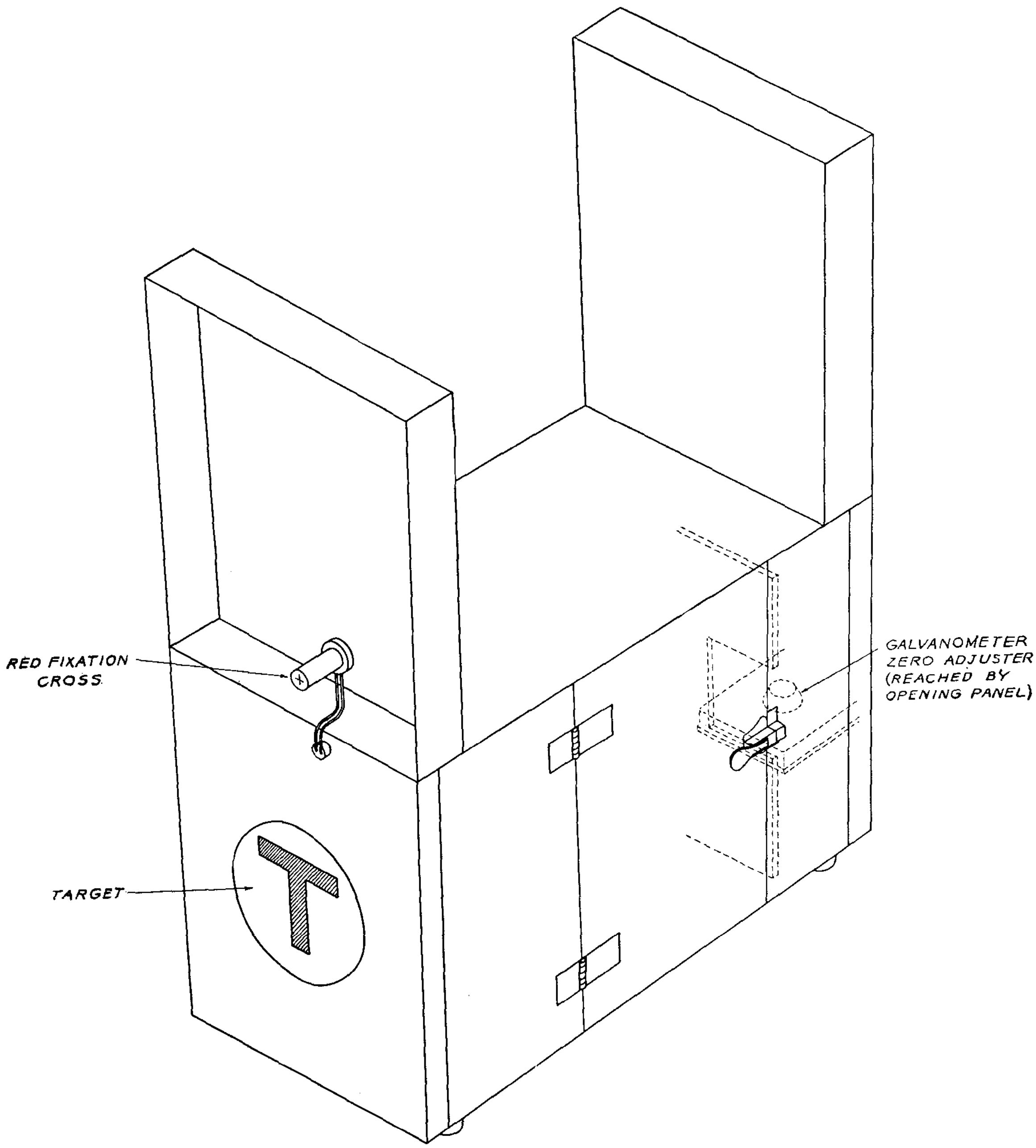


FIGURE 1. NDRC adaptometer, Model III, front view.

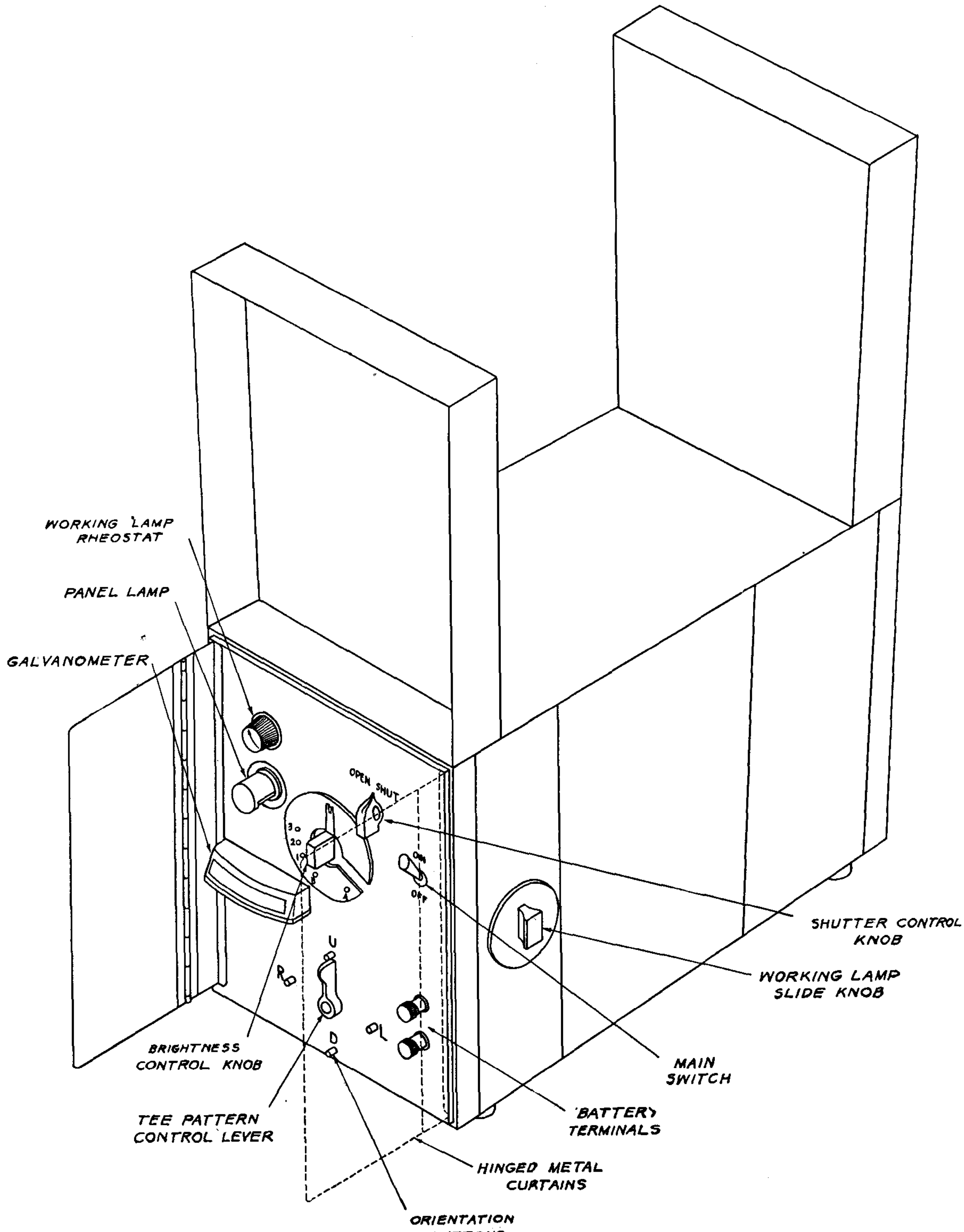


FIGURE 2. NDRC adaptometer, Model III, rear view.

[redacted]

demanded in the lookout situation. Eight men can be tested at one time, which makes it possible to increase the number of trials given each subject without increasing the total testing time. The adaptometer is simple in construction and operation and has easily controlled illumination. It has but two disadvantages: it requires a fairly large, dark room for testing and it is larger than most other available adaptometers.

9.4 EXPERIMENTS ON THE PREDICTION OF NIGHT LOOKOUT PERFORMANCE

Study of the validity of adaptometers and night vision tests was undertaken by the Applied Psychology Panel at the request of the Navy. The investigation was carried out by Project N-115 working in cooperation with the Medical Research Laboratory at the Submarine Base, New London, Connecticut.[2] Data were sought on the following points: (1) the reliability of a group of night vision tests, (2) the consistency of night lookout performance, or the reliability of measures of such performance, and (3) the relations between night vision test scores and night lookout performance.

9.4.1 Reliability of Night Vision Tests

The tests which were used in the experiment were the following:

1. The radium plaque adaptometer. This instrument was in use as the official U. S. Navy adaptometer. From a distance of five feet the subject reported the orientation of a T figure on a 5-degree white field of low illumination. The score was the number of correct observations.

2. The NDRC adaptometer, Model III. This test, as described in Section 9.2, was the same as the radium plaque adaptometer test so far as the man being tested was concerned. A slightly lower stimulus brightness level was used with this instrument than the level provided in the radium plaque adaptometer test.

3. The clockface adaptometer. This was a modified form of the method of testing night vision used by the amphibious forces. The subject was scored in terms of the distance from which he could correctly report the orientation of a white hand on a clockface which had white markers at 12, 3, 6, and 9 o'clock.

4. The Tufts-SDS test of night vision. This was a visual acuity test (Snellen E's) at low illumination.

5. A Purkinje test. This was a test of adaptation time, the time required for the subject to adapt sufficiently to see one of four test areas of low illumination.

6. The Hecht-Schlaer adaptometer, RCN Model. This instrument was the official adaptometer of the Royal Canadian Navy. It measured the subject's low brightness threshold for a 1/5 second flash.

7. A test of visual-motor coordination carried out at low illumination. This test was developed by K. U. Smith for the Committee on Aero-Medical Research.

8. The night lookout training stage. Performance on the training stage was scored in terms of the voltmeter setting necessary to see a ship model on the stage. Records were taken at the end of each of three training sessions.

9-12. Four personality tests.

13. Score on the verbal portion of the General Classification Test.

These tests were administered to a group of 150 seamen, none of whom had had experience as a lookout. Shortly after their arrival at the submarine base the men were given the tests listed above. The testing required three days. They were then given a regular three-day course of instruction in night lookout training. During the following three days the entire group was retested on the series of tests.

Data on reliability of the tests may be summarized by saying that none of the visual tests had a reliability coefficient high enough to be satisfactory for efficient selection. Test-retest correlations for four of the tests are shown in Table 1.

Reliability values for the other visual tests, where the distributions of test scores were particularly skewed, were obtained by the chi-square method. Four-cell tables were set up with various cutting scores applied to divide passing and failing groups on the first and

second tests. Values of chi-square significant at the 1 per cent level of confidence or better were obtained at some cutting level for all the tests. It was frequently true, however, that this confidence level was obtained only for cutting scores which would be impractical for use in Service selection; i.e., cutting scores which failed as many as 40 per cent of the total population.

TABLE 1. Reliability of night vision tests. ($N = 150$)

Test	Test-retest correlation
Clockface adaptometer (from Amphibious Forces)	+.27
Tufts-SDS test of night vision	+.19
Purkinje test (adaptation time)	+.53
Hecht-Schlaer adaptometer (Royal Canadian Navy Model)	+.42

9.4.2 Consistency of Night Lookout Performance

In the study of night lookout performance, 40 of the subjects were taken aboard ship each night and records were taken of their ability to pick up and report an approaching target vessel. Several target runs were made during each watch. A watch lasted 30 minutes to an hour and each man stood three watches. Target range and bearing information were secured by radar. Because of interference from airplane searchlights the data for only three nights of testing could be used. Several other factors combined to reduce the size of the test population still further. There were finally only 58 lookouts on whom useful performance scores were available.

In order to obtain a measure of the reliability of lookout performance, comparisons were made between the average range at which a particular man spotted a target on his first, third, fifth, etc., trials and the average range at which he spotted the target on his second, fourth, sixth, etc., trials. The product-moment correlation between the odd-trial average range and the even-trial average range for the 58 men was +.56. The performance was not very reliable. Reasons for expecting unreliability in lookout performance scores are many. They include trial-to-trial variation in target size, target speed, target angle, and sky brightness, the small number of trials, the relative inexperience of the men, and the small difference in ability found among the 58 men.

9.4.3 Relations between Night Vision Tests and Lookout Performance

Ten of the tests were useless in predicting the night lookout performance of the 58 men in the criterion group. These ten tests were the radium plaque adaptometer, the Purkinje test, the Hecht-Schlaer adaptometer, the Smith visual-motor coordination test, the night lookout training stage, the four personality tests and the verbal score on the General Classification Test.

The three remaining tests were positively related (according to chi-square measures) to night lookout performance. These tests were the NDRC adaptometer, the clockface adaptometer, and the Tufts-SDS test. Of these the clockface adaptometer and the Tufts-SDS test were valid only for cutting scores set high enough to fail more than 20 per cent of the group. They were, therefore, deemed impractical for use at the moment. The NDRC adaptometer, which differentiated between good and poor lookouts at a level which did not fail so many men, was at best a poor predictor of night lookout performance. It should be observed that although scores on this instrument were positively related to lookout performance, scores on the radium plaque adaptometer were not. Since these two instruments are very similar in nature and show a relation to one another which is nearly the same as the reliability of each, it is believed that the difference in predictive value between the two may have been due to the small number of men in the criterion group who failed the radium plaque adaptometer test and may, therefore, have been a chance difference.

In summary of this experiment, then, it was concluded that none of the seven tests of night vision and none of the six other tests tried

proved to be useful as methods of predicting the quality of shipboard night lookout performance.

Additional data on the efficiency of the radium plaque adaptometer in selecting night lookouts was obtained in another experiment[4] carried out as part of a shipboard study of night lookout performance which is described in detail in Chapter 6 of Volume 2, Applied Psychology Panel. Unfortunately, the radium plaque adaptometer does not yield scores that can be correlated easily with lookout performance scores. In the first place, it gives an all-or-none type of score: the man either passes or fails. In the second place, the man is tested at only one brightness level. Thus, between two men who pass or between two men who fail there may be very large differences in ability at the tested or at untested levels of illumination. It is possible, nevertheless, to make some rough comparisons between radium plaque adaptometer scores and the performance scores obtained aboard ship.

Of the 114 men who served as subjects in this experiment, 63 had been tested on the radium plaque adaptometer by one of the ship's pharmacist's mates, who was a qualified operator of the adaptometer. Of the 63, 50 had passed and 13 had failed the test. When the lookout performance of these men was measured in terms of ESF scores (equivalent square feet of target visible at 1,000 yards; see Chapter 6, Volume 2, Applied Psychology Panel for complete definition), it was found that the average score for those who had passed the radium plaque adaptometer test was 64.7 ESF, a score about equal to the average for all the 114 men. The average score for those who had failed the test was 43.9 ESF, a score which was better than the average score for the 114 men. Ten of the 13 men who failed the adaptometer test had better lookout scores than the mean score of those who passed the test.

TABLE 2. Comparison of radium plaque adaptometer scores for best and poorest lookouts.

Group	Mean lookout score	Radium plaque adaptometer record		
		Pass	Fail	No record
12 best men	12 ESF	8	1	3
12 poorest men	196 ESF	7	1	4

The data were considered in another way: the very best lookouts and the very poorest lookouts as measured by ESF scores were compared with respect to adaptometer performance. The results of this comparison are summarized in Table 2.

The difference in lookout scores between these two groups was clearly significant in a practical sense as well as in a statistical sense. The 12 best men could, on the average, see a ship about four times as far away as the 12 poorest men. But the two groups were equivalent with respect to radium plaque adaptometer performance. This comparison showed, as did the previous experiment, that the relation between scores on the radium plaque adaptometer and the performance of men as night lookouts is essentially zero. Recognition of this fact appeared in the "Manual of Procedures for U. S. Naval Classification Centers," NavPers 15082, dated February 1945, page 150, which contains these instructions:

> This test (Radium Plaque Adaptometer) is administered under the cognizance of the Bureau of Medicine and Surgery. The Bureau of Naval Personnel has directed that scores should not be used in determining qualifications for any school or rate, including Armed Guard. Scores for this test should not be recorded on the Enlisted Personnel Qualifications Card.

9.5 CONCLUSIONS ON LOOKOUT SELECTION

Failure to achieve satisfactory methods for selecting night lookouts hinges on at least three factors: the unreliability of night vision tests, the unreliability of lookout performance measures, and the probable importance for lookout performance of traits other than those measured by the tests used. Either a different kind of scotopic acuity is involved in the two sets of measurements (adaptometer and lookout situation), or the lookout scores are primarily measures of motivation and application

to the job rather than measures of visual ability. Only future research can determine which is the case.

Toward the development of more efficient lookout performance much can be done by increasing the prestige of the lookout, by the better training of lookouts and lookout officers, by formalizing the lookout organization aboard ship, and by making routine inspections of lookout procedures.[3] These matters are discussed in more detail in Chapter 6 of Volume 2, Applied Psychology Panel.

Chapter 10

SELECTING NAVY TELEPHONE TALKERS

By *Dael Wolfe* [a]

Summary

A SPEECH interview to rate the potential ability of men as shipboard telephone talkers was developed. The interview was conducted over Navy sound-powered telephones. It included the pronunciation of numbers and of all common American speech sounds, repetition of commands, and extemporaneous description. Men interviewed were rated as "well qualified," "qualified," or "not qualified."

A training course on the administration of the speech interview and the rating of speech quality was developed and taught to a number of classification interviewers.

The speech interview was adopted by the Bureau of Naval Personnel for routine use in the classification of enlisted personnel.

10.1 INTRODUCTION

A great deal of the communication within a Navy ship is by means of sound-powered telephones. The term "sound-powered" is used because the phones are activated by the energy of the speaker's voice without an external current source. These phones have the obvious advantage of allowing communication to be independent of the ship's power sources. But they cause difficulties in communication: fidelity is poor and the energy level is low. Recognizing these disadvantages, the Navy requested the Committee on Service Personnel in March 1943 to assist in the selection and training of those members of a ship's company who could talk most effectively over the sound-powered phones. The request was accepted and assigned as Project N-109 to the Psychological Corporation for solution. The work on *selection* of telephone talkers is described in this chapter; that on *training* of telephone talkers in Chapter 10 of Volume 2, Applied Psychology Panel.

The particular impetus for the project was a request for the immediate classification of men on the USS *New Jersey.* After work on this ship was finished, the speech interview as a means of rating telephone talkers was adopted at other Navy installations.

10.2 A SPEECH INTERVIEW FOR THE SELECTION OF TELEPHONE TALKERS

A speech interview[1] to rate the potential ability of men as shipboard telephone talkers was developed. During the interview the men pronounced numbers, repeated commands, and gave an extemporaneous description. All common American speech sounds were elicited during the interview.

Because of the urgency for immediate classification of men for the particular ship, the USS *New Jersey,* on which speech work began, it was impossible to take the time necessary to validate the speech interview as a selection method. Instead, methods of eliciting and rating speech were adapted from college speech-proficiency interviews. The interview was modified during preliminary exploratory work with 290 recruits selected at random from the Receiving Ship at the Philadelphia Navy Yard.

Officers concerned with classification had stipulated that the speech rating should divide the enlisted personnel into three categories as talkers: well qualified, qualified, and not qualified. The speech interview was intended to meet this requirement. It was included as a part of the classification process and preceded the general classification interview. In this way a man's speech rating was available to the interviewer when the man's duty assignment was decided upon.

In April 1943, project members began administering the speech interview to the prospective crew of the USS *New Jersey.* Speech ratings were obtained for 2,114 enlisted men

[a] This chapter is based upon the work of Project N-109.

on a seven-point scale: superior, very good, good, average, poor, very poor, and inferior. These ratings were immediately evaluated by the judge into the three categories requested by the classification officers. The testing program, in addition to supplying the data desired by the ship's officers, permitted the study of (1) the performance of skilled judges in making rating judgments, (2) the frequency of various types of speech deviations, and (3) the relationship of the several aspects of speech to intelligibility over voice communication systems.

While the testing program was in progress, 419 men who had been given the speech interview were also tested for listening and memory-span abilities by representatives of the Harvard Psycho-Acoustic Laboratory. An investigation of the relation between listening ability, memory span, and speech ability was thus made possible (STR of NDRC Division 17, Volume 3).

The following conclusions appeared to be warranted.[1]

1. A speech rating can be assigned to each member of a ship's enlisted personnel through a speech interview at the time that he is classified for ship's duty.
2. The original seven-point speech ratings fell into an essentially normal distribution.
3. Neither sea experience, telephone-talking experience, nor education is closely related to an enlisted man's speaking ability as judged by this method.
4. Speech traits that accompany limited intelligibility are: poor articulation, foreign or regional dialect, too weak a voice, poor control of intensity, and a voice quality heard as harsh, hoarse, or husky.
5. All factors considered, those speaking the general American dialect were judged to be slightly more intelligible than those speaking the other two major dialects, southern or eastern.
6. Expert judges making independent ratings agree closely in evaluating intelligibility.
7. Expert judges agree more than 90 per cent of the time in placing men in one of the three categories of well qualified, qualified, or disqualified.
8. Speaking ability is not highly related to memory span or listening ability.

10.3 TRAINING OF SPEECH INTERVIEWERS[3, 4, 5, 6, 7]

Results obtained in the selection of telephone talkers for a number of individual ships led to a demand for wider use of the speech interview. Expert speech men were not numerous enough to interview all recruits. It therefore became necessary, first, to determine whether or not the speech interview could be given successfully by classification interviewers[2] and, second, to develop a course for training classification interviewers to administer the speech interview.

Project members prepared a short manual on the selection of telephone talkers.[4] This manual, and a college textbook on voice and articulation (Fairbanks: *Voice and Articulation Drillbook*), were used as texts in a course given by project personnel to train classification interviewers to administer the speech interview.[5]

In addition to their reading, future interviewers also engaged in a number of types of drill. They were encouraged to practice speaking exercises during their free time. Each student was provided with five-minute recordings of his speech at the outset of the course and at the end of the course. Two Mirrophones were available all the time for practice. Class periods included lectures on speech; Mirrophone drill; practice in judging recruits' speech in face-to-face situations, from recordings, over telephones, and over ship's dial telephones; and practice in making repeated judgments of merit of recorded speech interviews.

Results obtained with this course indicated that classification specialists could be trained to administer the speech interview satisfactorily. The student subjects improved in their ability to evaluate speech, and as the training progressed their judgment coincided more and more closely with that of expert judges and with the average of group judgments. The results were regarded as tentative, as it was felt that more effective training methods might be developed in the future.

Typical correlations of the ratings made by petty officers trained to administer the speech interview with ratings of the same men made by an experienced speech teacher were about

.80. This level of agreement was considered sufficiently high to justify using classification petty officers for routine administration of the speech interview.

10.4 LIMITATIONS ON THE USE OF THE SPEECH INTERVIEW

The ultimate fitness of a telephone talker depends, not only upon the intelligibility of his speech, but also upon a fairly specialized knowledge of equipment, organization, and procedures, as well as upon his attitudes, sense of responsibility, and behavior under stress. Thus a rating of "qualified" or "well qualified" on the classification card should not be interpreted to mean that a man has been certified as fully capable of performing all the duties of a telephone talker. It means merely that the man's speech intelligibility is passable; it is such that with proper training and experience he can become a qualified or well qualified talker.

10.5 NAVY ADOPTION OF THE SPEECH INTERVIEW

As a result of the development of the speech interview and the studies in the training of classification specialists to administer it, the Navy began to use the method at Naval Operating Base, Norfolk, Virginia. Over 20,000 men were rated as telephone talkers at that station, and the ratings were noted on their qualifications cards. The speech interview was later approved as a standard classification procedure by the Bureau of Naval Personnel (letter, Pers-6372-MS-4, RS-F-4d, dated April 15, 1944). It became a routine part of the Navy classification and selection practice in a number of stations on both the east and west coasts.[8]

Chapter 11

PRINCIPLES AND DEVICES FOR MILITARY CLASSIFICATION

By *John L. Kennedy* [a]

Summary

COOPERATIVE WORK by the Navy and NDRC developed mass classification procedures for the crew of the USS *New Jersey*, for the Amphibious Training Command of the U. S. Atlantic Fleet, and for the crews of destroyers being manned on the west coast. The principles and procedures developed were adopted for use throughout the Fleet.

The objectives of classification are to assign each man to the job he can do best and to divide available manpower equitably among the various specialties and among the various units of a military force.

Requirements for the success of a classification program are described. A detailed knowledge of the requirements of each military specialty is necessary. Tests used to determine a man's qualifications should be as independent of each other as possible; they must be validated. Classification data must be kept up to date and available; procedures must be standardized; and the programs should be under the supervision of classification specialists.

Several devices to aid in securing reliable information for use in classification were developed and installed on several ships. A mechanical device which does much of the work of combining and weighting classification data is described.

11.1 INTRODUCTION

World War II demonstrated that the concept of an untrained citizen army, springing to arms overnight in a national emergency, is no longer tenable. Wars are fought more and more by specialists. Even the infantryman, the traditional "cannon fodder," emerged from the war as a specialist in the use of a complicated array of weapons requiring a considerable period of training to master. The large number of new weapons put further stress on specialization in the military forces and further complicated the training program. The growing specialization and the consequent time required for training in World War II emphasized the importance of scientific classification, allocation, and assignment of manpower.

Discussion of these topics is divided between this and the two following chapters. This chapter gives a brief account of the development of mass classification procedures by the Navy and NDRC and then presents some general principles of good classification procedure and some devices to aid classification. Chapters 12 and 13 present in greater detail improvements which could be made in recruit and advanced classification programs.

11.2 NAVY-NDRC DEVELOPMENT OF MASS CLASSIFICATION PROCEDURES

Early in 1943 the Applied Psychology Panel was requested to assist in the classification and assignment of the crew of the USS *New Jersey*. Growing out of this cooperative Navy-NDRC effort came two requests, (1) to develop classification procedures for the Amphibious Training Command, U. S. Atlantic Fleet, and (2) to aid in the development of classification procedures for men assigned to duty on new destroyers that were being commissioned on the west coast.

Simultaneously the Bureau of Naval Personnel was extending and improving its work in the classification field.

The success of these programs led to the adoption on a fleet-wide basis of the principles and procedures developed and tested in earlier programs.

[a] This chapter is based upon the experience of the Applied Psychology Panel and the work of several of its projects which cooperated with the Bureau of Naval Personnel in establishing classification procedures.

11.2.1 The USS *New Jersey*

Approximately 2,600 men made up the crew of the USS *New Jersey*, the first of this country's 45,000-ton battleships. Two-thirds of the men had never been to sea before their assignment to the *New Jersey*. The need to get the ship ready for fleet duty as quickly as possible was great. To expedite that process, the Bureau of Naval Personnel and NDRC were requested to analyze the nature of the ship's many duties, to test and interview each man, and to recommend an assignment for each which he could learn most rapidly and in which he could work most efficiently.

In addition to recommending assignment for each member of the ship's company, a mechanical method for the rapid identification of personnel with special skills and aptitudes was installed on the *New Jersey*.

Classification of the crew of the *New Jersey* was an emergency effort, done with little time for detailed preparation and without opportunity for validating all the tests used. But it was a successful effort. Fewer transfers to new duty occurred on the *New Jersey* than were necessary on other ships. Men learned their duties more rapidly. Fewer disciplinary problems arose. The Commanding Officer summarized these advantages in reporting: "The initial placement of personnel has definitely contributed to the apparently extraordinary rapid progress made during the shakedown period by the crew of this vessel." (Letter BB62/P16-3, Serial 570, August 31, 1943, to Chief of Naval Personnel).

11.2.2 Atlantic Fleet Amphibious Training Command

The second request for assistance in developing classification procedures for large groups of men came in a request from the Training Officer of the Amphibious Training Command, U. S. Atlantic Fleet, to assist in establishing a model classification program and to carry out research to improve the classification procedures already set up. Project N-117 with the State College of Pennsylvania was activated by the Applied Psychology Panel in order to meet this request. The project, in cooperation with officers from the Bureau of Naval Personnel and officers from the Amphibious Training Command, was given the opportunity to work out broad, general procedures for the allocation of manpower as well as the methods for classification into specific job assignments.

The Amphibious Training Program was a new development made necessary by the conditions under which World War II was fought. No blueprints existed for amphibious organizations; even the type of craft used was a matter of continuous development. The variety of specialized jobs to be done on board these craft meant that considerable time could be saved by the application of scientific procedures for handling the thousands of inexperienced officers and men who passed through the training program. The intensive work of the project and its collaborators resulted in the adoption of a "Manual of Classification Procedures for Amphibious Training Bases"[13] at all the amphibious training bases on the east coast.

Emphasis should be given the importance of the work in the Amphibious Training Command as it bears on the overall problem of mass allocation, classification, and assignment of personnel. A thoroughgoing attempt was made in this work to control the input to the training program and in the assignment process to "balance" crews in terms of abilities and aptitudes. In other words, instead of leaving the "raw material" for the classification and assignment process to chance, planned use was made of test scores, previous job history, special skills and potentialities (determined by testing and interviewing *before* entrance into the training program) so that the output for assignment to specific ships could be controlled. In the assignment to ships after training at the amphibious training bases it was discovered that an enterprising officer was able to round up a crew which consisted entirely of "top" men. This meant that the next crew had to be made up from a pool with the top removed by the previous assignment. Such a process, when continued, resulted in marked unevenness in the ability of various crews.

From the standpoint of the officer in charge of a particular ship, it was obviously desirable

to have as good a crew as possible. But there were not enough good men to provide first-class crews for every ship. From the standpoint of the Service as a whole it was desirable to divide the good men among the ships, to "balance" the crews. One or two incompetent crews could have a disastrous effect on the efficiency of an entire attacking force if they failed to work properly in an attack landing.

The principle of "balancing" and apportioning available brains and ability fairly among the crews had its initial full-scale trial in the work of this group.

11.2.3 Classification of Destroyer Crews

Project N-116a with Stanford University was activated as the result of a request from the Commander, Fleet Operational Training Command, Pacific, to assist in developing adequate techniques for classification for gunnery billets on all new destroyers manned on the west coast.[11] The selectometer, a mechanical aid in classification and assignment of large groups of men, was one contribution of this project. It will be described in Section 11.4.4. Its development and use illustrate a second principle of mass classification, namely, the utilization of mechanical sorting, weighting, and summing procedures to make possible the rapid and standardized utilization of information about personal characteristics in classification and assignment.

11.2.4 Fleet-wide Adoption of Mass Classification Procedures

The Commander-in-Chief, United States Fleet, basing his directive on the success of the programs described above, on February 16, 1945 directed the Chief of the Bureau of Naval Personnel to develop a similar classification program for fleet-wide application (second endorsement to ComPhibTraLant ltr. FE 25/A7-3, Serial 0043, dated January 24, 1945).

11.3 GENERAL PRINCIPLES OF GOOD CLASSIFICATION

The success of a military force depends upon the ability and training of the units of which it is composed. The force as a whole will normally be most successful if the units are balanced in terms of ability and training. There are, therefore, two primary objectives of a classification program. One is to assign each man to the duty in which he can be most useful. The other is to distribute the available supply of ability evenly among the units composing a force. The classification experience of World War II indicates that realization of these two objectives requires observance of the following general principles.

1. Detailed knowledge of requirements for each military specialty is necessary.
2. Aptitude measures should be independent.
3. Aptitude tests must be validated.
4. Classification data should be constantly available.
5. Classification procedures should be standardized.
6. Classification should be supervised by classification specialists.

These six general principles are discussed in the following sections.

11.3.1 Detailed Knowledge of Requirements for Each Military Specialty Is Necessary

Determining the requirements of a military specialty requires an analysis of the job itself by the techniques described in Applied Psychology Panel, Volume II, Chapter 14, in order to secure a detailed account of what is done. This analysis makes it possible to determine the mental and physical qualifications necessary to perform the duties of the job.

It is sometimes fairly easy to determine the necessary qualifications. For example, the fact that the distance separating the eyepieces of a heightfinder can be adjusted through a fixed and limited range determines one necessary characteristic of a heightfinder operator: his interpupillary distance must fall within the range of possible instrument settings.

More frequently it is necessary to determine empirically the exact requirements of a job. Such an empirical study is desirable both because it provides information on essential

qualifications and because it avoids the waste involved in setting too high selection requirements. If selection standards are set too high, time is wasted in locating men who qualify, and the men selected are made unavailable for assignment to duties in which their high qualifications would be more valuable.

It is not difficult to select a group of men who will be successful in the Signal Corps, in Naval Aviation, or in some other specialty. It is more difficult, and in the long run more important, to determine the needs of each specialty, to assign each man to the specialty in which he will do best, and thus to divide the able men equitably among the various specialties.

11.3.2 Aptitude Measures Should Be Independent

If aptitude tests which purportedly measure different aspects of ability show high correlations with one another, they are measuring some general type of ability and failing to do the job for which they were intended; they will select men of generally high ability rather than picking out those high in the particular abilities needed in a given job. For example, if mechanical ability is important in one job and verbal ability in another, men cannot be selected efficiently for the two jobs unless the mechanical and verbal tests are relatively independent of each other. Aptitude measures should be as independent of each other as possible.

11.3.3 Aptitude Tests Must Be Validated

No aptitude test can be used with complete confidence until it has been validated in accordance with the procedures described in Chapters 14 and 15. In wartime, tests must sometimes be used before validation is complete. It is then necessary to rely on the judgment of a test expert in deciding whether or not to use a test for classification purposes. The test expert may be wrong, as may an engineer; some tests have been worthless, and some bridges have collapsed. But when validation data are not available, the test expert's judgment is the best available basis for deciding what tests to use. The U. S. Navy Basic Classification Test Battery, for example, was put into use before being validated. Later evidence showed it a valid battery and confirmed the judgment of the officers who decided upon its immediate adoption.

11.3.4 Classification Data Should Be Constantly Available

Classification data must be available in order to be used. The statement is so self-evidently true that its inclusion here should not be necessary. But there are so many ways in which classification data, carefully secured at one time, may become unavailable when needed that the principle demands emphasis.[6]

Classification data may become unavailable through any of these common means:

1. Failure to enter test scores and similar information on the man's record card (WD AGO Form 20 in the Army; Q-card in the Navy; or their successors).
2. Failure to provide for recording school grades or other measures of demonstrated proficiency.
3. Failure to transfer record cards when a man is reassigned.
4. Failure to provide proper mechanical devices for storing, sorting, and using record cards.
5. Failure on the part of an officer in charge to understand the nature and use of classification data supplied to him so that records are not properly preserved and hence are not there when wanted.

These failures can all be avoided by proper indoctrination of officers and by inspection checks to determine that personnel policies are complied with.

11.3.5 Classification Procedures Should Be Standardized

Many classification interviewers and personnel assignment officers are required to keep the classification machinery operating. Their work must be done in accordance with established

[redacted]

and standardized forms, procedures, and rules. If all officers and men responsible for classification were highly competent experts in their tasks, rules could be fewer; but since they are not and in a large military organization never will be, standard procedures are necessary.

Standardized procedures may result in some misassignment of men, but for the great bulk of the men being classified the total amount of misassignment will be smaller than it would be if each interviewer had complete freedom of choice in making his recommendations. The selectometer (Section 11.4.4) is one example of mechanical standardization of the task of recommending assignments. The superiority of test scores to interviewers' judgments in selecting men for school training (Section 1.5.3) illustrates another aspect of the advantages of standardized classification regulations.

There will of course always be some men whose qualifications are so unusual, and some jobs for which the requirements are so unusual, that they must be treated individually. These exceptions in no way minimize the importance of the general principle.

11.3.6 Classification Should Be Supervised by Classification Specialists

Personnel classification is a technical specialty. The specialists differ from the amateurs in making fewer mistakes and in knowing how many mistakes they make. It may appear easy to examine a man's record card, to note that he worked as a garage mechanic, and to decide that he would probably be successful as a machinist's mate. It is more difficult to determine which other qualities are involved in determining whether the garage mechanic will or will not be a good machinist's mate. And it is also more difficult to determine which men who have not had mechanical experience will be good machinist's mates. The classification specialist knows how to solve these more difficult problems, and he knows just how successfully they can be solved with present methods. Personnel classification is a technical specialty. It requires technically trained officers to ensure successful classification.

11.4 DEVICES TO AID MILITARY CLASSIFICATION

11.4.1 Tests of Aptitude and Achievement

Aptitude and achievement tests bear a large share of the burden of determining the particular specialty to which each man will be assigned. Specific aptitude tests have been described in Chapters 2 to 10. Their construction, standardization, and validation are discussed in Chapters 14 and 15. Achievement tests are discussed in Chapter 17 of Volume 2, Applied Psychology Panel.

11.4.2 Interview Aids

Several devices to aid the interviewer in securing relevant information and to help in assigning appropriate weight to the various items of information were developed.

A Point-Score Method for Evaluating Naval Personnel

Interview forms listing specific elements of experience considered important for success on destroyers were prepared. A weight was assigned to each item in terms of its importance as judged by a number of individuals well acquainted with the duties to be performed. The total score for any individual could then be determined by summing the weights of the items. The weighting of interview information was by this means made more standardized.[7]

Interviewing Guides

Another aid to the standardization of interviewing procedures was the development of interviewing guides for destroyer billets.[5, 12]

Work-Readiness Tests

Work was begun in 1945 on the development of short series of questions patterned after the U. S. Employment Service oral trade questions. They were intended for use in determining a man's readiness (in terms of knowledge) to assume the duties of a particular billet. Only one, for distilling plant operators, was actually constructed.[9, 10]

[redacted]

11.4.3 Shipboard Classification Devices

In processing men through shore-based classification centers a great deal of information is obtained about each man's experience, ability, and potential usefulness aboard ship. There is still, however, a problem of making that information readily available aboard ship. Emergencies arise and casualty replacements must be made from within the ship's company. In order to make these replacements quickly and effectively it is desirable to have some method of determining which members of the crew possess the needed skills or experience.

The most efficient solution of this problem consists of having the pertinent data concerning each man entered on a single record card. Various companies, the Findex and the McBee for example, have developed simple mechanical methods of sorting these cards so that the cards for men who possess the desired characteristics can be quickly separated from the rest.

These systems possess the distinct advantage of making all the recorded information about a man immediately available. There is no forgetting.

On three occasions during World War II the Applied Psychology Panel was asked to help in devising mechanical sorting systems for shipboard use. Each of the three systems is briefly described below.

Findex Equipment for the USS *New Jersey*

Following the classification of the crew of the USS *New Jersey* (Section 11.2.1), Findex equipment was installed aboard ship to aid in the rapid location of men needed in the operation, particularly in the emergency operation, of the ship.[1] The manual contained detailed instructions to supplement those furnished by the Findex Company.

Sorting the Q-Card

Members of the staffs of Applied Psychology Panel Project N-106 and of the Classification and Selection Section, Bureau of Naval Personnel, cooperated in devising a sorting scheme to increase the usefulness of the U. S. Navy Enlisted Personnel Qualifications Card (Q-card). The scheme made provision for coding seven items of information (Rate, Branch, Occupational Skills, Gunnery Experience, Foreign Language Fluency, Rate Designator, and Talker Qualification) along the two perforated edges of the Qualifications Card. An accompanying manual[3] described the details of the coding and illustrated how various types of sorting jobs could be done once the cards had been coded.

Using templates, the guide holes on the edge of the card can be punched according to a code system which provides for coding the following seven items of information along the two perforated edges of the Q-card.

1. Branch (seaman, artificer, etc.—eight categories in all).
2. Rate (100 categories in all).
3. Special Occupational Skills (142 categories).
4. Gunnery Experience (two categories).
5. Foreign Language Fluency (two categories).
6. Rate Designator (two categories).
7. Talker Qualification (two categories).

The branch code number and punch code numbers illustrate the coding system:

Code and Punch Code Numbers Used in Recording Branch

Branch	Code Number	Branch Code Number
Seaman	1	SF and 1
Artificer	2	SF and 2
Engine Room	3	2 and 1
Aviation	4	SF and 4
Special	5	4 and 1
Commissary	6	4 and 2
Steward	7	SF and 7
Specialists	8	7 and 1

A second example shows the punch code numbers for several job titles of the 142 categories of special occupational skill which can be coded on the Q-card.

The scheme preserves the existing format and content of the Q-card. It is comparatively easy to use, and the cost of the four pieces of equipment which it demands is nominal. The system itself does not impose any definite limitation on the size of the groups of cards dealt with, but it is believed that it can be used most conveniently with groups of less than 500.

Experimental Shipboard Personnel Record System

The Enlisted Classification and the Test and Research Section of the Bureau of Naval Personnel, Applied Psychology Panel Project

[redacted]

N-106, and the McBee Company prepared a Personnel Record File for installation aboard the USS *Catoctin*.[8]

Table of Special Occupational Skill Codes (Alphabetical)

Code No.	Job Title	Punch Code Number Hundreds	Tens	Units
1	Accountant	0	0	SF-1
22	Accounting Clerk	0	SF-2	SF-2
2	Actor	0	0	SF-2
64	Advertising Manager	0	4-2	SF-4
77-a	Aerographer	0	SF-7	SF-7
109	Agent, Purchasing	1	0	7-2
36	Air Conditioning and Refrigerating Engineer	0	2-1	4-2
3-a	Aircraft Metal Worker	0	0	2-1
102-a	Aircraft Pilot	1	0	SF-2

The face of the card contains room for recording 53 items of information descriptive of a man's aptitudes, experience, and training. A selection of 30 of these items can be coded along the edges of the card where they are accessible for sorting purposes.

Use of Shipboard Filing Systems

The record of each man is reduced to a single card of, say, 8 by 10½ inches. These cards are filed alphabetically, making it an easy matter to locate information on the experience, training, and special abilities of any individual man aboard ship.

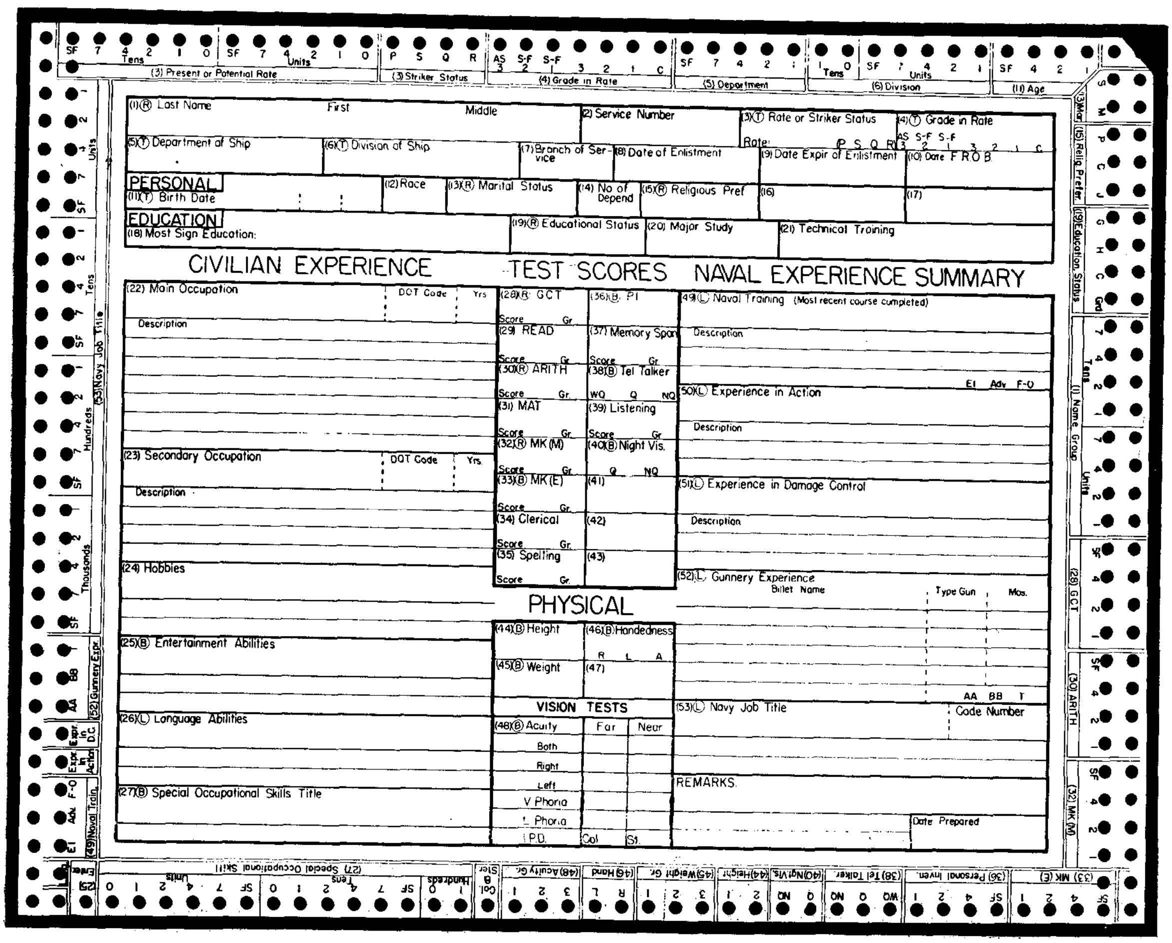


FIGURE 1. Record card for shipboard filing system.

The record card used was of the keysort type. It is shown in Figure 1 in front view. The back of the card contains room for a chronological record of the man's naval training and experience.

When a need arises for men of specified qualifications, simple mechanical sorting can quickly locate them. As examples, one could pick out all men rated as qualified telephone talkers, all men having a particular occupational skill, all

men who have been under fire, all experienced in damage control, or all left-handed men. If combinations of skills are desired, it is possible to select on several traits. One could, for example, locate in a few minutes every qualified telephone talker who had been under fire and who had had experience in damage control.

The usefulness of such a record system depends upon the thoroughness with which information is entered on the record card and the care with which the records are kept up to date. If those things are done properly, the possession of such a record system makes it possible for the ship's officers to locate the men they want when they want them.

11.4.4 Mechanical Weighting of Classification Data

The most novel development of the Applied Psychology Panel's work in the field of classification procedures was the invention of a device called the selectometer which automatically weights and combines a man's scores (or other measures) on each of the factors considered important in selecting men for a particular job.[4] The selectometer was developed by Project N-116, Stanford University, in cooperation with officers of the Operational Training Command, U. S. Navy Pacific Fleet.

The basic aim of classification is to select the men best suited for each of a number of positions. This selection is made by considering the qualifications of the individuals available for placement and comparing those qualifications with the requirements of each job. This is not easy. Remembering and weighting, more or less simultaneously, qualities which the classifier believes are related to success in different jobs presents an extremely difficult task. In actual practice, the interviewers will frequently overlook some factors, will be inconsistent in the importance attached to the several factors, and will, in fact, seldom know how they weight the factors which they do consider in making an assignment.

One way to introduce uniformity into this procedure is to weight the factors and sum the weighted scores mechanically. There are various ways to do this part of the classification job. It can be done by hand computation or with

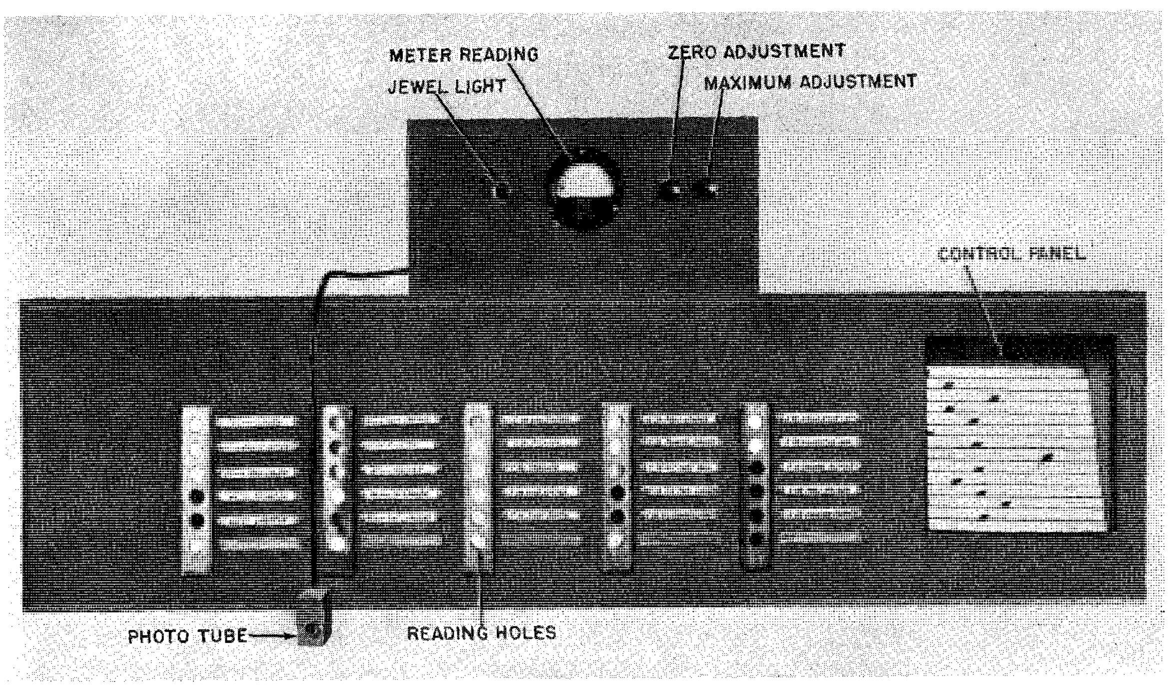


FIGURE 2. The optical selectometer.

[redacted]

the aid of punched cards. The selectometer does the same job more rapidly.

THE OPTICAL SELECTOMETER

Operation and Construction. Figure 2 is a photograph of the optical selectometer as it appears ready for operation. When a man is to be classified, data taken from his Q-card are set into the control panel by moving pointers to the proper settings. For example, if he is six feet tall the "Height" pointer is moved to the mark "71-72." When all the other variables (age, test scores, etc.) are similarly treated, the instrument is considered "set up" for the individual.

The extent to which the man is qualified for various jobs is then determined from the "reading holes." Each hole represents one job, and the amount of light streaming through each hole determines the degree of qualification for the corresponding job. The intensity of light is measured by a phototube connected to a microammeter through an amplifying circuit. The meter reading thus provides a quantitative measure of fitness for each position.

The instrument uses an optical-mechanical method of weighting qualifications. The extent of qualification for each job is indicated by the intensity of the beam of light from a fluorescent light source through the corresponding hole in the front panel. The beam of light is reduced in intensity by interposing filters in proportion to the lack of qualifications which an individual possesses. Thus, the more qualified a man is for a job, the more intense is the beam of light through that job hole. The less qualified he is, the greater is the reduction by means of filters, and when an individual possesses a characteristic which would disqualify him for a job, the use of an opaque filter eliminates all light.

The filters are mounted on plates which slide behind the front panel. When a plate is moved to the position which represents a particular score or class interval of a variable, the holes in it are so lined up as to bring the proper filter directly behind the front reading hole. Each of the plates represents one variable. The total amount of light transmitted by the set of filters indicates the degree of qualification for the job.

To offer a standardized method of representing the degree of qualification of any category, a set of filters was prepared from 35 mm motion picture film of varying density or light transmission value. For convenience, the filters

TABLE 1. Filters used for weighting.

Filter No. (weight)	Shade	Desired percentage of light transmission	Interpretation for degree of job qualification
0	Clear	100	Best for job, or no adverse relationship
1	Very light	90	Slightly less desirable, or probably not quite the best
2	Light	80	Fairly desirable, or may be only slight handicap
3	Light gray	70	Probably adequate, but likelihood of some handicap
4	Medium gray	60	Possible success, but probably some drawback
5	Dark gray	50	Doubtful; considerable probability of failure
6	Dark	40	Unlikely to succeed, chances more definitely against
7	Very dark	20	Bare possibility of success; great risk
8	Opaque	0	Reject; very high probability of failure

were given numbers from 0 to 8 according to the density. Table 1 indicates the amount of light transmitted by each filter and its interpretation in terms of degree of job qualification.

The weights assigned to each category within each of the variables used in making gun station assignments are given in reference 4. These weights were assigned on the basis of analyses of the duties and requirements of the gunnery jobs for which assignments were to be made, a study of the characteristics of the men available for assignment,[2] and conferences with a number of gunnery officers to refine and standardize tentatively assigned weights.

The method of weighting used in the selectometer combines the advantages of the multiple cutting score technique with those of the multiple correlation method. The selectom-

[redacted]

eter, like the cutting score procedure, permits one to eliminate a man on the basis of one disqualifying characteristic alone, even though his scores on other variables may be extremely good. On the other hand, when a very high qualifying score has a higher predictive value than a score just above the cutting line, the selectometer, like the regression equation, gives extra weight to high scores; it thus takes full advantage of any predictive differences in the scores above the cutting point.

In its application at the Precommissioning Training Center, Treasure Island, California, to classification for destroyer gun stations, the following variables were incorporated in the instrument: GCT, Arithmetic, Mechanical Knowledge, Height, Near Vision, Far Vision, Age, Years of Education, Interest (for a type of gun), Talker Test, Build, Sports in Which Qualified, and Leadership. In its use at other centers some different variables were used according to the measures which were available in the classification office.

The list of variables used in the selectometer does not include all the factors which are important for success in gun stations. For example, a man's success in these positions depends upon his motivation, dependability, work habits, willingness to follow orders, ability to get along with others, his morale, his ability to stand up under fire, and a host of other factors. The reason for not including such variables is simply that no adequate measures or estimates of them are obtainable. There is a lack of adequate methods of making such estimates and a lack of personnel and time to obtain such data as might be available.

An interesting feature of the selectometer is that the unsatisfactory and unreliable appraisal of some traits becomes immediately apparent when their inclusion in the list of selectometer variables is considered. For example, interviewers who think they have been considering a man's "leadership" in recommending his best assignment find themselves facing a new problem when they are asked exactly how much weight should be given to what degrees of leadership. When more reliable methods of estimating such traits become available and when agreement on their importance is reached, they may readily be incorporated into the selectometer system.

The important variables, rate and experience, were omitted from the selectometer because most officers are already accustomed to using rate and experience in making gun station assignments. They would be doubly weighted if they were included in the selectometer. Furthermore, there is a definite difference in policy among officers with respect to assignment of rated men to gun stations.

Reliability of Selectometer Ratings. The reliability of the selectometer method depends not only upon the consistency of the instrument itself but upon the accuracy of the operators. An experiment to determine the reliability of the method was undertaken with two experienced operators. The selectometer raw scores for 101 individuals were obtained on two occasions a day or two apart. Product-moment correlations were obtained for four typical positions: 5″/38 sight-setter, .84; 20 mm gunner, .96; 40 mm second and third loader, .97; 20 mm trunnion operator, .90. Analysis of the scores on which there was variation seemed to indicate that inaccuracies on the part of the operators constituted the major source of unreliability. From all the present evidence it appears that the consistency of the instrument itself is extremely high.

The accuracy of the light filter method was determined by comparing optical selectometer scores with those obtained by the punched card method described below. For 20 mm gunners ($N = 188$) the correlation was .90. When corrections for errors of operators are considered, one may conclude that the method of using light filters for combining weights is highly accurate.

The ultimate test of the validity of the selectometer is the overall comparison of selectometer scores with gun station proficiency at sea, especially during combat. Practical considerations made this test of validity impossible.

The Electrical Selectometer

Figure 3 is a photograph of a demonstration model of the electrical selectometer. In this instrument four variables are used. They are

[redacted]

height, far vision, age, and GCT, as represented by the switches on the front panel. Each stop of the switch represents a category of the variables as indicated in the previous discussion.

The job switch at the right-hand side of the panel includes a sample of jobs with different requirements. These are 20 mm gunner, 40 mm gun captain, 40 mm pointer, 5″/38 first loader, 5″/38 passer.

istic is judged to have for the job. This is, in essence, the opposite of the optical selectometer, where the current is decreased in proportion to the penalty. In the electrical selectometer, a resistor is analogous to the filter in the optical selectometer; a large resistor which permits very little current to pass through is comparable to a light filter which reduces the light only a small amount, and a very low resistor transmitting a large amount of current has the

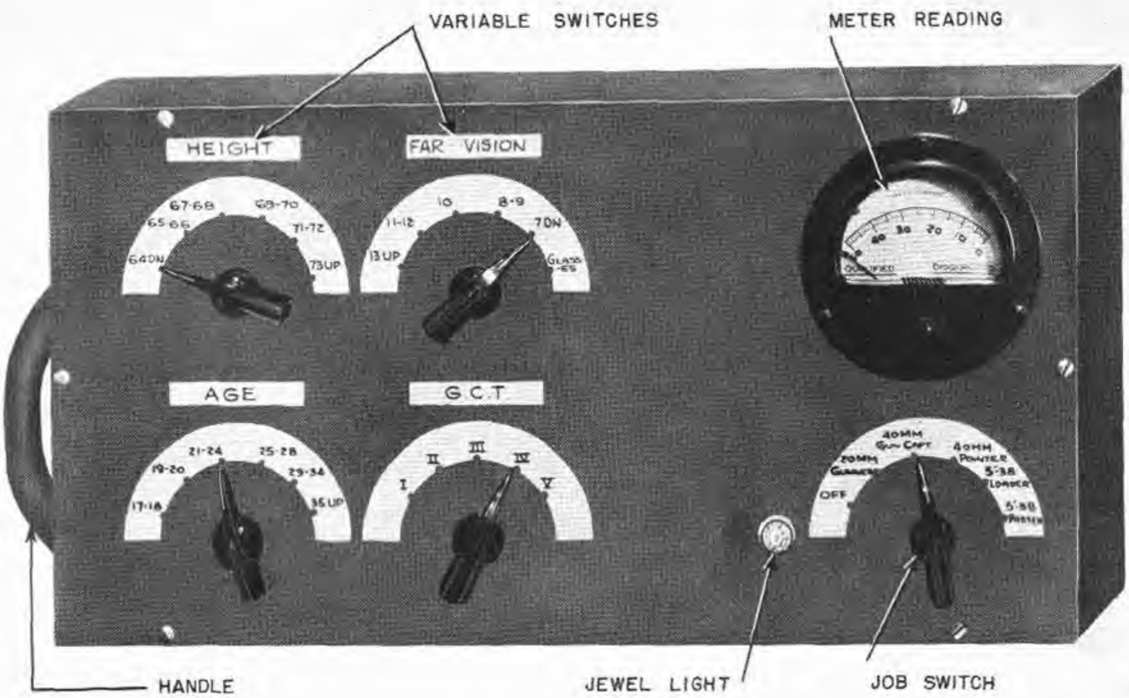


FIGURE 3. The electrical selectometer.

same function as a very dark filter. In other words, in the electrical selectometer, no current through the ammeter means "Fully Qualified," while maximum current through the meter means "Disqualified." In order to avoid confusion the scale on the meter was reversed, so that 50 means qualified, and 0 disqualified.

When an individual's potentialities for various jobs are to be determined, the variable switches shown in Figure 3 are set for his characteristics. Other variables may be incorporated as methods of estimating are developed. The job switch is then turned to each job in succession, and the scores are read directly from the meter. These scores may be used in the same fashion as those obtained from the optical selectometer.

In the electrical selectometer, current flowing through the meter is increased in proportion to the size of penalty which a character-

The wiring of the instrument is shown in Figure 4. For the sake of clarity, only three jobs are shown. When the switches are set for a variable, a conductor of a certain resistance is thrown into the job circuit, thus adding an amount of current proportional to the esti-

[redacted]

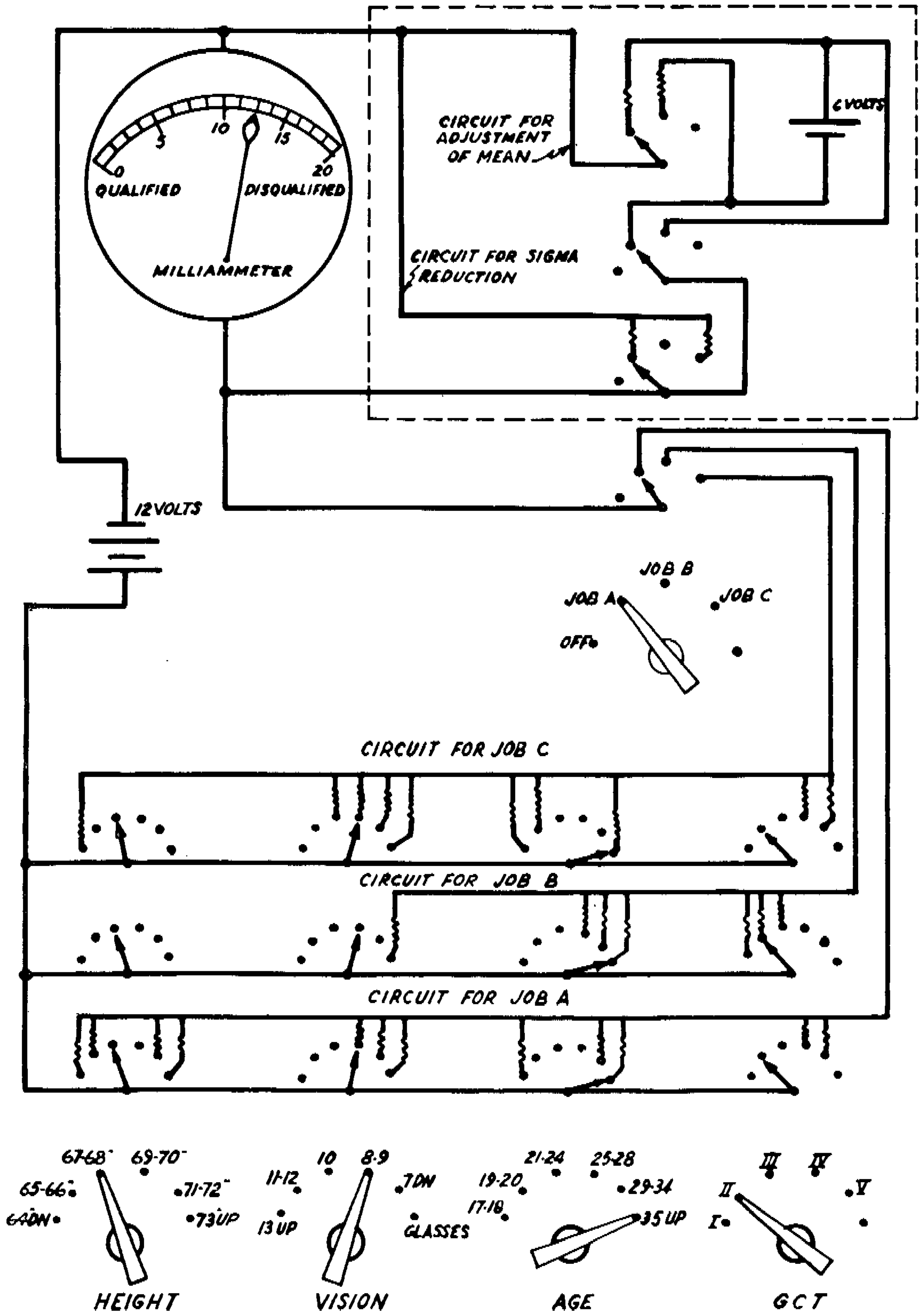


FIGURE 4. Wiring diagram for optical selectometer.

[redacted]

mated degree of disqualification. The source of current is a 12-volt battery, but ordinary 110-volt alternating current through a rectifier tube could replace batteries. An ammeter reading from 0 to 50 milliamperes is used to measure the current flowing in the circuit.

By means of a special circuit, as indicated within dotted lines in Figure 4, the distributions may be changed so that the limits for group scores will be more nearly similar for all jobs. The wiring diagram also shows how, by the introduction of another power source, current may be added to or subtracted from the job circuit so that all the scores on one job may be lowered or raised by any desired constant amount. By using the shunt resistance across the meter at the same time, the sensitivity of the meter is reduced so that the standard deviation of these scores is lowered. If the distributions were all normal, these circuits would make it possible to obtain exact standard scores directly.

The original design of both optical and electrical selectometers gave the same weight to a category regardless of the scores on other variables. In certain special instances, it may be found desirable to make the weight assigned to variable A depend upon the value of variable B. In other words, it may be decided that the weights should not be assigned to either variable independently but should always be assigned to a particular combination of scores on the two variables. The method may be extended to deal with special combinations of more than two variables if necessary. The selectometer may, in fact, be constructed to give a certain designated score to any combination of any number of variables.[4] This procedure will not only permit the incorporation in the selectometer of advantages claimed for the profile comparison method, but it will standardize the procedure in a far more complete fashion than a more subjective method ever allows.

Advantages of the Electrical Design over the Optical Design

1. The electrical instrument is easier to construct; there are fewer mechanical problems to be worked out; the circuit is merely a compound, not a complex, one.
2. The instrument may be built more compactly than the optical model.
3. The use of standardized commercial resistors eliminates the problem of having special filters developed. These resistors may be obtained commercially with a 5 per cent tolerance or may easily be selected for greater accuracy if so desired.
4. The use of additional switches would permit incorporation into the instrument of all the advantages of the "profile" method in any desired fashion.
5. The use of circuits to adjust the mean and standard deviation of the distributions would permit direct reading of scores in either absolute or relative terms.
6. The electrical instrument is easier to keep in adjustment because of its relative mechanical and electrical simplicity.

Punched Card Weighting of Selection Variables

Since punched cards are in extensive use in many classification centers in the Navy, a procedure was devised to apply this technique to the problems of the kind which stimulated the development of the selectometer. The use of punched cards as described below yields raw scores which correlate highly with scores obtained from the present instrument. As indicated earlier, the correlation was .90 for a 20 mm gunner.

With the punched card method, the same weight numbers may be used as with the optical selectometer. But the numerical values assigned to these weights are as follows:

Weight number	Numerical value
0	00
1	03
2	07
3	12
4	18
5	25
6	34
7	54
8	99

For each category of every variable, a group of cards is gang-punched with the numerical value of the weights assigned to this category

[redacted]

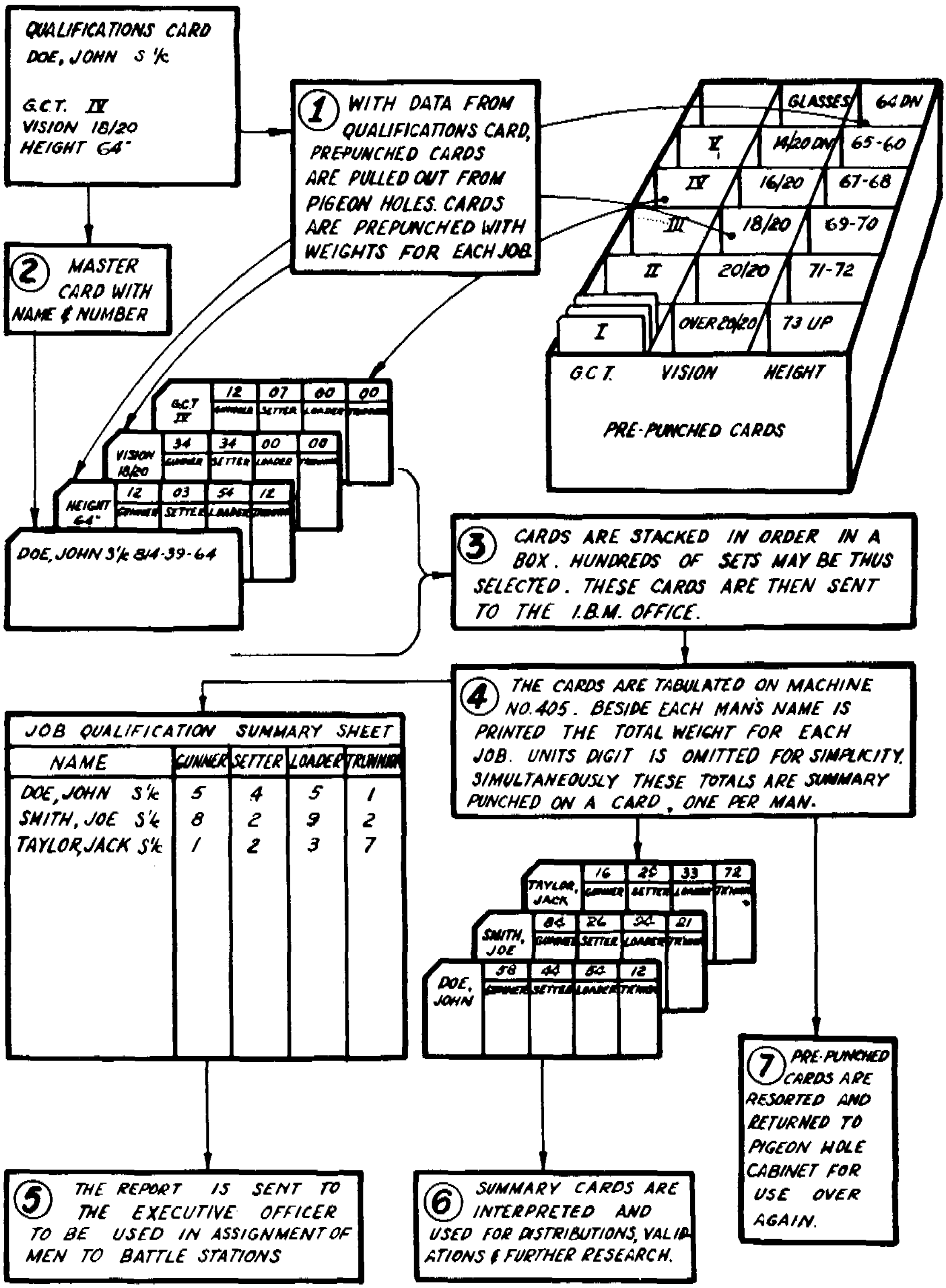


FIGURE 5. Graphic outline of punched card method of weighting selection variables.

for each job. The same two columns are assigned to each job on all the cards for all categories of all variables so that the sum of the numbers in any set of two columns will give the total score for a job. These prepunched cards are filed in a pigeonhole cabinet, each hole containing the cards for one category.

In Figure 5 is presented in graphic form an outline of the general method to be followed in processing an individual through this classifying procedure.

Evaluation of the Selectometer

On the basis of the results obtained with the optical selectometer and of the evidence available from the development of the electrical selectometer and the punched card method, an evaluation of the application of selectometer procedures to classification problems may be made. Whether one uses an optical or an electrical model or uses punched cards is a matter of convenience. The principles and advantages are the same. The chief points are:

1. The selectometer focuses attention on the importance of carefully developed job specifications. In order to be incorporated into the selectometer, job specifications must be stated very explicitly. This precise statement facilitates the validation of billet specifications.

2. The selectometer standardizes the use of billet specifications. By combining and weighting scores mechanically, it eliminates any individual differences in the importance attached to various characteristics.

Various weighting techniques for combining the variables may be employed. Once the weighting technique has been agreed upon, the selectometer will follow it uniformly for all men.

3. The selectometer provides a method of obtaining qualification scores of each man for a number of different jobs. The time required to do this task for a large number of men and jobs without some partly mechanized system would be prohibitive.

4. Different classification needs can be accommodated by revising weights, adding or subtracting variables, or raising or lowering cutoff scores.

While the selectometer was designed and initially used for making gun station assignments it appears to be equally applicable to the problem of making school assignments for recruits or, indeed, to help in any classification situation in which a large number of men are to be sorted into a relatively large number of separate groups on the basis of a number of different measures.

[redacted]

Chapter 12

ORGANIZATION OF A CLASSIFICATION PROGRAM FOR RECRUITS

By *Norman Frederiksen*[a]

Summary

The procedures followed during World War II in classifying and assigning enlisted personnel in the Navy are briefly described.

Improved classification would result from:

1. Using a primary test battery to separate the men of school quality from those who should be assigned to general detail, and a secondary test battery to determine the particular school to which each man in the first group should be assigned.

2. Improving the interview by standardizing the weighting assigned to various factors on which the interviewer's recommendation is based.

3. Improving the basis for filling school quotas by establishing assignment pools and by better determination of proper priorities in filling quotas.

4. Continuous research to keep classification tests and methods abreast of changing military requirements.

5. Indoctrinating regular officers in modern methods of personnel classification.

12.1 OBJECTIVES AND PROBLEMS OF RECRUIT CLASSIFICATION

The classification officer at an Army Reception Center or Naval Training Center faces a difficult problem during wartime. Every week thousands of recruits are brought to the center. They represent very nearly the maximum possible range of variation with respect to almost any possible measure except age and physical condition. Some are doctors of philosophy, while others have never attended school; some are geniuses or potential geniuses, while others are feeble-minded; some are highly skilled mechanics, engineers, and instrument makers. while others are poets, violinists, and artists. Within a period of a few days the classification officer and his staff are required to evaluate the actual or potential abilities of each recruit and assign him to that type of military training or duty where he will make his greatest contribution toward the military prosecution of the war.

In peacetime there may be more time available to the classification officer, and the weekly flow of men may be smaller, but the problems of proper classification remain. For the efficiency of a military force depends upon the efficiency of the men who compose it. Classifying and assigning each man to the job that he can do best will make the whole force an effective one.

The objective of military classification is, then, to make the most effective use possible of manpower. This objective may be accomplished by evaluating the aptitudes, abilities, skills, interests, educational and occupational experience, and physical and personal characteristics of each man and on the basis of these evaluations assigning the man to a particular type of military training or duty. (Examples of the tests used for classification and evidence of their value were given in Chapters 2 to 10.) In addition, the aim is to make the optimum assignment of each man in the *minimum amount of time*. Given sufficient time, men might eventually "shake down" or gravitate to the billets for which they are best qualified, even without careful classification procedures; but it is important, particularly in wartime, that each man be given a satisfactory assignment as quickly as possible. The saving of time in making suitable assignments, especially in the early stages of a war, may in effect increase considerably the manpower available for combat.

More specifically, the objectives of classification of recruits are to answer such questions as the following:

[a] This chapter is based chiefly on the experience of NDRC Project N-106 in helping to develop the Navy's classification tests and procedures.

1. Who (among the thousands of recruits to be classified in a particular week) already possess sufficient knowledge, skill, and qualities of leadership to warrant their being sent immediately to their duties as technicians or noncommissioned officers?

2. Who possess the aptitudes and intellectual qualifications which warrant sending them to a Service school for training in some military specialty?

3. Of those who are judged to have the aptitudes which warrant specialist Service school training, who shall be trained as radarmen, who as gunners, who as quartermasters, who as tank drivers, who as radio mechanics, etc.?

4. Who lack the qualities which would justify investing time and money in special technical training and therefore should be assigned to general, nontechnical duty?

These are samples of the questions which the classification officer must answer. The job of classifying the recruits is made difficult, not only by the fact that the men are coming continuously in tremendous numbers and must be processed within a short period of time, but also by various other considerations. For one thing, a large proportion of the men have, because they are young, no background of civilian occupational experience which would help in making assignments; therefore aptitudes, or ability to *acquire* specific skills, must be evaluated *in advance* of training. A conscript army in peacetime would be composed of young recruits without significant occupational experience. In the case of older men, occupational experience may be a factor which can be used in classification; but even then its usefulness is limited, since civilian occupations do not bear a one-to-one correspondence to military jobs. Few if any civilian jobs are similar in nature to those of a military gunner or torpedoman. Some civilian occupations, such as insurance salesmen and textile workers, are available in numbers far greater than needed by the Services. Furthermore, assignments must sometimes be made to schools where, because of the rapidity of development of new military techniques, the aptitudes required and even the curriculum taught are not fully known. In short, large numbers of men are available whose skills and qualifications are not obviously apparent; they must be fitted to a great variety of military jobs, some of which require rare or unusual aptitudes, or aptitudes which are difficult to evaluate, or even aptitudes the nature of which is not fully known.

12.2 CLASSIFICATION PROCEDURES IN USE BY THE NAVY

The classification procedures currently in used by the Navy[5,6] are described briefly in this section. Treatment of Army classification procedures is not included because almost all the work of the Applied Psychology Panel on recruit classification was done in the Navy.

12.2.1 Testing and Interviewing

The recruit is given a physical examination (including psychiatric screening) by the Medical Department soon after his arrival at the station. The report is forwarded to the Classification Department. In the days which immediately follow, the recruit spends most of his time in activities sponsored by the Selection Department: (1) he takes the tests of the U. S. Navy Basic Classification Test Battery (see Chapter 2); (2) he listens to lectures and sees movies calculated to inform him about the schools and rates available to enlisted men; and (3) he fills out an "Aid to the Interview Blank" on which he describes his educational and occupational history and gives other information which the classification interviewer may later use in making his classification.

Preparation of Record Forms

In the meantime, the preparation of two important personnel record forms is begun. The first of these is the Classification Department Electric Accounting Punched Card. In it is punched information which the Personnel Department supplies—Service number, name, rate and grade, company number, etc. As soon as the answer sheets for the Basic Battery tests have been scored (by means of the test-scoring machine), the scores are also punched in the card.

The second personnel record is the Enlisted

Personnel Qualifications Card (Q-card), which, when processing is complete, becomes a part of the recruit's permanent Service record which accompanies him throughout his service in the Navy. The Q-card is printed on continuous form stock for use with an electric accounting machine. By using this machine, the data so far punched in the Electric Accounting Card are printed on the top two lines of the Q-card. Then additional data from the Medical Department report and the names of particular schools or duties for which the recruit is qualified or disqualified are entered manually.

The Interview

The recruit is now ready for the interview, which is conducted by a Specialist (C). Specialists (C) are petty officers who have been selected because of special qualifications (such as experience with the U. S. Employment Service) to serve as classification interviewers. They are given a special course of training prior to classification work, as well as in-Service training. The interviewer has at hand the Aid to the Interview Blank as previously filled out by the recruit, and the Q-card which is partially processed as described above. During the interview the Specialist (C) adds to the Q-card additional information pertaining to educational and occupational history, hobbies, and the like.

12.2.2 Recommending Assignments

Now comes the most important function of the Specialist (C). He makes recommendations, on the basis of all the information so far obtained, as to the assignment of the recruit. The success or failure of the classification procedure hinges largely on the wisdom of these recommendations.

The interviewer ordinarily makes two recommendations on the basis of the recruit's test scores, his occupational experience,[b] physical qualifications, interests, etc. The first recommendation represents the most appropriate assignment, in the opinion of the Specialist (C), and the second recommendation represents the next most appropriate assignment. The recommendations may be for assignment to general detail, to an elementary Service school, or to an operational school.

In addition to making specific recommendations, a Quality Classification Code is assigned to each recruit by the Specialist (C); this code is supposed to be a qualitative rating in terms of the recommended assignment. Code *1* is used to designate men judged to be exceptionally well qualified for the recommended school assignment. Code *2* men are, in general, as well qualified as Code *1* men but lack a "definite pattern of qualification" and therefore can be assigned to a school *similar* to that of the first recommendation. Code *3* men do not have the high qualifications of the above groups for the recommended assignment but are sufficiently well qualified for assignment if needed. In addition, codes may be assigned which indicate recommendations that the man be rated, that he be held for reassignment, or that he be reinterviewed, perhaps by a psychiatrist. The classification of the man may also be postponed until special tests can be given; this is done in the case of men tentatively selected for a certain school on the basis of Q-card data. Code *0* indicates lack of qualifications for school assignment; this quality classification always accompanies a recommendation of assignment to general detail.

The recommendations and Quality Classification Code are recorded by the Specialist (C) on the Q-card, and later punched into the Electric Accounting Card. These punched cards are now ready for sorting.

12.2.3 Filling Quotas

As soon as possible, a summary of the recommendations is sent by teletype to the Enlisted Distribution Division of the Bureau of Naval Personnel. At the Bureau, quotas are made up, on the basis of fleet demands and the number of men recommended for Service school at each of

[b] A booklet called *United States Navy Occupational Conversion Tables* (October 17, 1942) has been prepared for the purpose of aiding the interviewer in evaluating civilian work experience. This booklet gives the Navy rates which are related to a large number of the job titles in the Dictionary of Occupational Titles of the U. S. Employment Service.

[redacted]

the various naval training centers. Then the Bureau sends to the Selection Department its quotas—the number of men who must be sent to each of the various duties and training activities.

The Order of Assignment

It is, of course, not likely that for a given week the distribution of Service school quotas will match the distribution of recommended school assignments. A specified order of assignment which may be applied in making assignments to any Service school has therefore been developed. In filling quotas, the cards are sorted and men selected according to the following order:

1. 1st recommendation, quality class 1
2. 1st recommendation, quality class 2
3. 2nd recommendation, quality class 1
4. 2nd recommendation, quality class 2
5. 1st recommendation group, quality class 1
6. 1st recommendation group, quality class 2
7. 2nd recommendation group, quality class 1
8. 2nd recommendation group, quality class 2
9. 1st recommendation, quality class 3
10. 2nd recommendation, quality class 3
11. 1st recommendation group, quality class 3
12. 2nd recommendation group, quality class 3

If a quota for quartermaster school, for example, were to be filled, the cards would be arranged according to the order of assignment, and the quality class "1" men whose first recommendation is quartermaster would be chosen. If sufficient cards are found to fill the quota, the job is finished so far as quartermaster school is concerned, and one can go on to another school's quota. But if there are not enough quality class "1" men whose first recommendation is quartermaster, it is necessary to make further selections; quality class "2" men recommended for quartermaster school would next be picked out, and so on. By "first recommendation group" is meant a group of schools with selection requirements similar to those of the school recommended *first;* for example, if the quartermaster quota had not been filled after choosing men in the first four categories of the order of assignment, the next step would be to sort out quality class "1" men recommended for schools in the same group as quartermaster—aerographer's mate or Specialist (Y) control tower operator schools.

Completing the Personnel Records

After the cards have been sorted to fill all quotas, the selection lists are printed from the cards by means of the electric accounting machine. The assignment and date of transfer are punched in the cards, which are then duplicated. One card is filed by the Classification Department and the duplicate goes with the man to his Service school assignment; these duplicate cards are later used for research purposes. The Q-card is photographed and then inserted in the man's service record; the film is sent to the Navy Department.

12.3 SUGGESTIONS FOR IMPROVEMENT OF THE NAVY'S CLASSIFICATION PROCEDURES

A good deal of the research[4] on recruit classification has been based on the punched cards which were mentioned above. This research, together with firsthand observation of procedures and talks with classification officers, has led to a number of suggestions for improvement of the Navy classification procedures. Some of the more general suggestions are discussed in the following section.

12.3.1 Use of Primary and Secondary Test Batteries

The first suggestion for improvement of the classification program is based on the difficulties that are inherent in using one basic battery of aptitude tests for all the various purposes involved in selection. Use of the same tests for such diverse purposes as separating school from nonschool material and separating

[redacted]

candidates for fire control school from candidates for torpedo school is likely to result in tests which are not entirely suitable for either purpose. Furthermore, administration of the entire battery to the large group of men who will not go to Service school is wasteful of testing and test-scoring time.

It is possible to administer a greater variety of tests without increasing the total time spent in test administration by the device of using a short *primary* battery of tests for the purpose of selecting the school material, and using *secondary* test batteries for the purpose of assigning men to specific Service schools.

The details of such a program as is here suggested would of course have to be modified in accordance with research findings; the following suggestions are tentative.

The Primary Test Battery

The primary battery, which would be administered to all recruits, might consist of only two tests: a purely verbal test, somewhat like the present General Classification Test, and a mechanical test, which might consist of items resembling the pictorial items of the present Mechanical Knowledge Test. This second test might be entirely a picture test, in order to lower its correlation with the verbal test as much as possible. Item analysis techniques should be used, in developing these tests, to ensure as low a correlation as possible between the verbal and the mechanical tests. This primary battery of two tests could then be used to separate those men who are unlikely to succeed in any type of Service school from those who are likely to succeed in some Service school. (The cutting scores, selection ratios, or other devices would of course have to be determined on the basis of research investigations.)

The primary battery would be useful, not only for identifying the men who are the best risks so far as Service school success is concerned, but also to classify these men into groups depending upon the general *type* of Service school for which they should be considered. The scheme might work somewhat as follows: men low on both verbal and mechanical tests should be assigned to general detail; men who are high on verbal but low on mechanical should be considered as candidates for a "clerical" type school (yeoman, signalman, radioman, etc.); men who are high on mechanical but low on verbal should be considered as candidates for a "mechanical" type of school (gunners mate, basic engineering, electrical, metalsmith, etc.); and men who are high in *both* should be considered as candidates for the more highly technical schools (fire control, radio matériel, etc.).

The Secondary Test Batteries

The next step in classification would be to administer to the men in each group a secondary battery of tests especially designed to aid in assignments to schools within that group; for example, the men in the clerical groups might be given a battery of tests including a clerical test, a spelling test, a radio code aptitude test, and a blinker aptitude test. The men in other groups would similarly be given batteries of tests appropriately selected to aid in assignment to the schools of the group.

Administering a primary and a secondary test battery would take no longer than the present method of administering every test to every man. Some men would take only the primary battery; the others would, in general, take no more tests than they do at present. This system would probably result in more valid tests, since each could be designed for a specific purpose. It also admits the possibility of greater use of individual tests, because, by subdividing the number of men to whom a secondary test battery must be given is considerably reduced. The objection to the procedure is that problems of scheduling are involved. It would no longer be possible to order an entire company to take a single battery of tests. Methods would have to be evolved for assigning men to take tests individually instead of by companies.

12.3.2 Improvement of the Interview

A second suggestion for the improvement of the classification procedure arises from cer-

[redacted]

tain researches which seem to indicate that assignment to Service schools on the basis of test scores alone would result in somewhat more successful classification than assignment on the basis of interviewers' recommendations.[2, 3] The interviewers, it seems, tend to weight too heavily the information they themselves obtain, at the expense of the aptitude test scores. Fairly rigid requirements with respect to use of test scores should, accordingly, be set up, and the interviewers should be required to follow these standards. Furthermore, these studies suggest the importance of further study of the validity of the other items of information on the Q-card, and if the results justify it there should be set up specific rules for weighting each item of information in making the school recommendation.

A standardized procedure for evaluating the various factors that are significant in making an assignment was developed by one of the Applied Psychology Panel projects.[1] The procedure, called a point-score method, involves the assignment of uniform weights to carefully selected and explicitly stated factors. Evaluations based on such a procedure would certainly be far more reliable and objective than the ordinary interview evaluation. The validity of the procedure would depend upon the care with which the factors and their weights are selected. The objection to the method is that the factors and weightings would vary from one assignment to another and points would need to be assigned separately for every assignment considered.

12.3.3 Improvement of Procedures Involved in Filling Quotas

The third suggestion pertains to the quota system of allocating recruits to various activities. As was described above, the fact that quota distributions fail to match the recommended school assignment distribution has necessitated the use of a complex Order of Assignment by which the sortings progressively take men less and less well qualified for the assignment. In practice, not only were all twelve categories of the order of assignment used, but it was sometimes necessary to go beyond this order and take men recommended for general detail or for schools entirely outside the Order of Assignment categories.[2]

Improving Procedures for Determining Quotas

The situation could be vastly improved by two courses of action. The first and most fundamental of these steps is to improve the procedures for setting up the quotas themselves. (This is, of course, not a responsibility of the Classification Department.) In the past it has sometimes appeared that quotas were determined in a somewhat whimsical fashion; the quota for a given school might suddenly double or triple in size, while a large quota for another school might suddenly disappear entirely. More adequate canvassing to determine fleet needs well in advance, together with careful study of the summaries of recommendations supplied by local selection officers, would go a long way toward making it possible to assign most men to the training for which they have been more specifically recommended.

The problem of priorities in filling quotas needs careful consideration. At present little or no data are available on the quality of men needed for various types of duty; yet definite priorities have in effect been established. The "Eddy Test" for pre-radio matériel school candidates is administered at the time of induction, thus making it possible to make an early selection of only very high quality men for these schools. Fairly high cutting scores on Basic Battery tests have been set for all enlisted personnel given air training assignments. Men for the amphibious forces and armed guard, on the other hand, are assigned after the best men have been chosen for other schools, and those left tend to be of rather low caliber. This establishment of priorities in classification no doubt has some justification; but no data as yet exist which indicate whether or not men of high quality are needed for radio matériel more than, for example, fire control or that they are needed more by the air arm than for surface ships. It would seem that every branch of the Service needs at least some

men of high quality; although the proportion may vary, competent men are needed to provide leadership in each branch. Research is needed to answer such questions and eventually to determine more precisely how men with various degrees of competence should be distributed.

Establishing an Assignment Pool

A second course of action which would help in ensuring the assignment of more men to their recommended duty would be the establishment of a pool of men whose classification has been completed and who are awaiting assignment. The present practice at most centers is to assign the men classified each week on the basis of that week's quota demands. Delaying assignment only one week would go a long way toward making it possible to compensate for fluctuations in quotas. Such a pool was established in an unofficial way at one center, with good results; it would have been done at other centers but for the lack of space in which to house the men. With the existence of a pool from which to draw and with quotas which are made up with due consideration of long-term fleet needs and available men, it should be possible to select all men on the basis of the first half of the Order of Assignment and thus to place a much larger proportion of the men in duties for which they are recommended.

A pool of men from which quotas can be filled could easily be established merely by making men available for assignment any time during, say, the last three weeks of basic training. The flexibility needed for filling quotas would thus be accomplished without the necessity of additional housing facilities. The time of the men in the "pool" would be occupied by training while awaiting assignment. The only disadvantage is that the period of basic training might be shortened somewhat for some of the men.

12.3.4 The Need for Research

The need for research on classification procedures has been mentioned several times in the foregoing paragraphs. In order to keep selection procedures abreast of changes in methods and techniques of warfare, a permanent organization for conducting a continuing program of research is needed.

12.3.5 Indoctrination of Officers

In order for the values of a good classification procedure to be fully utilized and to be of maximum service in efficient utilization of military manpower, it is necessary that officers outside as well as in the Classification Department have some understanding of classification procedures and sympathy with the objectives of classification. The consequence of lack of understanding is revealed by an extreme case where the classification of a large group of men had been completed and the assignments reported to the Personnel Department. The Personnel Department, however, ignored the assignment lists and made their own assignments on the purely arbitrary basis of alphabetization of names. The indoctrination of such a large and varied group of men as the officers in the United States Navy is, of course, a difficult task, especially in the hurry of wartime training. A few weeks' training in classification and personnel procedures might be of very little value when given along with navigation, gunnery, communications, and a host of other important subjects in an indoctrination school. In peacetime, however, it would be possible to train the majority of the future officers of the Navy who would become the key men in the event of another war. This training can and should be started at the U. S. Naval Academy at Annapolis.

Chapter 13

ORGANIZATION OF AN ADVANCED CLASSIFICATION PROGRAM

By *Norman Frederiksen*[a]

Summary

THE METHODS used by the Navy in advanced classification and reclassification of enlisted personnel are briefly described.

Several Applied Psychology Panel projects have contributed procedures or recommendations for the improvement of advanced classification. The principal ones are:

1. More extensive construction and use of objective methods of determining the actual proficiency with which a man can perform the duties of a particular billet. A number of examples of performance tests developed for this purpose are described.
2. Improvement in the methods of rating shipboard proficiency. There are situations in which performance tests cannot be used but in which ratings by superiors can be obtained. Methods of obtaining more reliable ratings are given.
3. Use of brief oral tests, called work readiness tests, which can be administered by a classification interviewer to determine how much a man knows about the details of a particular type of duty.
4. The indoctrination of line officers in the methods and information already available which they could use in making shipboard reassignments.
5. Improved methods for coding and filing the personnel data available for the men assigned to a particular ship or unit make those data more easily available and more likely to be used.

13.1 OBJECTIVES AND PROBLEMS OF ADVANCED CLASSIFICATION

The objectives, problems, and techniques of advanced classification are in general similar to those of recruit classification, which was discussed in the preceding chapter. Both are concerned with the assignment of men, on the basis of actual or potential abilities, to those types of military training or duty where they can make their greatest contribution to the winning of a war. Both are concerned with achieving the optimum assignment for each man in the minimum amount of time.

The differences, however, are sufficiently great to justify their separate discussion. The procedures involved in the classification of recruits are concerned primarily with the evaluation of a man's *potentialities for training,* since the majority of recruits have had no training or experience directly related to Navy jobs; advanced classification procedures, on the other hand, more often involve the *evaluation of training and experience in the Navy* with respect to a specific duty or billet. Assignments in recruit classification are based upon information rather hurriedly obtained through techniques of aptitude testing and interviewing, while advanced classification assignments can be based to a greater extent on actual proficiency in Navy jobs, as reflected in records of performance in a training situation, in a billet aboard ship, or on tests of proficiency. Advanced classification, furthermore, should involve the balancing of crews so that all ships and departments have their fair share of superior men as well as those of average or inferior ability.

Advanced classification is at once more complex and more simple than recruit classification. It is more complex in that the specific assignments that can be made are more varied and depend upon a greater variety of measures. They are simpler in the fact that assignments can be based more often on measures of proficiency in a given task, rather than on *predictions* of success from measures of aptitude.

Advanced classification may be thought of as including any classification procedures that

[a] This chapter summarizes the contributions and recommendations of a number of Applied Psychology Panel projects.

occur other than the original assignment of the recruit to some duty or training. Since a man may be reclassified a number of times during his career in the Navy, advanced classification is seen to be an activity which pervades all parts of the Navy, both ashore and afloat, and which is of even greater magnitude than the task of recruit classification. Some typical questions which might be answered by advanced classification are as follows:

1. Who among the members of the crew of a DD are qualified to use sound-powered telephones?
2. What battle station should be assigned to each crew member?
3. Who from a pool of available men should be assigned to each billet on a ship which is about to be commissioned?
4. What men should be recommended for striker billets?
5. What transfers between ships should be effected in order to obtain equitable distribution of abilities?
6. What men need special operational or advanced training to prepare them for assignment to the departments, divisions, or battle stations aboard ship for which they are recommended?
7. What men should be assigned to ships' company billets at shore stations?
8. What men need refresher training to keep them up to high standards of proficiency in their billets?
9. What men should be selected for officer candidate training (such as the V-12 program)?
10. What members of the crew have useful occupational skills or special abilities (interpreters, speedboat coxswains, cartographers, mechanical draftsmen, barbers, movie operators, entertainers) which are not indicated by rate, and how can these men be picked out when needed?
11. What men meet the minimum standards of visual acuity and night vision for duty as night lookouts?

These questions are representative of those which must be answered in advanced classification activities. The procedures currently in use by the Navy for supplying the answers to such questions are described in the following section.

13.2 ADVANCED CLASSIFICATION PROCEDURES IN THE NAVY[21]

Advanced classification centers are located at receiving stations, training stations, Navy yards, operating bases, and other centers where personnel may be assigned to specific duties. The staff of the classification center processes all men who pass through the station to which the classification center is attached, and also, in some cases, men attached to other nearby shore stations or ships. The processing includes the administration of tests, interviewing, bringing the Enlisted Personnel Qualifications Cards up to date, and making specific recommendations as to the assignment of each man.

13.2.1 Test Administration

The general policy with respect to test administration is to make sure that the record of test scores on the Enlisted Personnel Qualifications Card (Q-card) is complete. If Basic Battery test scores are already recorded on a man's card, the tests are not readministered (unless there is evidence that the scores are in error). According to directive, if any of the scores are lacking, the appropriate tests should be given and the scores entered. In addition to Basic Battery tests, various other group tests may be given when appropriate, for instance the Radio Technician Selection Test, the Winchman and Hatchman Test, Oral Trade Questions, and the Literacy Test. A number of special apparatus tests are also available, such as the Sonar Pitch-Memory Test, the Telephone Talker Test, and various visual tests for measuring such variables as acuity, phoria, depth perception, and color discrimination.

13.2.2 Interviewing and Recommending Assignments

One important function of the interviewer is to bring up to date the Q-card, adding any missing information which should have been recorded when the card was first prepared and

filling in the more recent history of naval experience and training. Cards which are obsolete or inaccurate are replaced. On the basis of the Q-card information, including test scores and records of experience and training, the interviewer then recommends assignment of the man to some duty or billet. The interviewer's general evaluation of the man's performance is also taken into account. The type of assignment will depend in part on the particular situation—if a man is a member of a ship's crew which is being processed, the recommendations would be in terms of that ship's watch, quarter, and station bill. The recommendation is made by matching the qualifications possessed by the man to the requirements of a particular billet; this can be done well only if the interviewer is well acquainted with the requirements for each billet. For example, he must know whether or not use of sound-powered phones is involved in a particular assignment; if so, the man assigned to that billet must be rated as a qualified telephone talker as well as meet the other specifications for the billet.

In addition to recommending billet assignments for each man who is processed, the advanced classification center may perform other services. Recommendations may be made for striker or petty officer billets; a recommended battle bill may be prepared in which specific general quarters stations are recommended; cross index files of potential strikers may be prepared (arranged according to both names and shipboard ratings); and cross-index files of special skills may be developed.

13.2.3 Indoctrination of Ship Personnel

The classification center furthermore attempts to make provision for the continuation of the classification program aboard ship, through the indoctrination of ships' officers as well as by providing cross index files for use aboard ship. In particular, one officer from each ship, designated by the commanding officer to serve as shipboard classification officer, is given special training in the use of the available data for selecting strikers, assigning billets and battle stations, and effecting transfers among divisions. A shipboard classification yeoman is also trained to process Q-cards, administer and score tests, and keep up-to-date cross index files and files of Q-cards.

The above is a brief account of advanced classification procedures. The various advanced classification centers actually differ somewhat in their mode of operation, because of local variations in equipment, personnel, and problems. Some are more effective than others because of better cooperation with other related departments. Some have been particularly successful in furthering the benefits of classification through indoctrination of shipboard officers, while others have failed to take steps to ensure the continuation of the program aboard ship.

13.3 SUGGESTIONS FOR IMPROVEMENT OF ADVANCED CLASSIFICATION

The suggestions for improvement of advanced classification procedures which are contained in this section are not new ideas; in many cases, the officers in charge of classification are fully aware of the problems and have already taken appropriate steps to correct the undesirable situations. However, the Applied Psychology Panel has made contributions in cooperation with classification officers; some of these contributions are discussed below.

13.3.1 Use of Proficiency Measures

As was stated earlier, advanced classification has the advantage, which recruit classification does not, that the men to be assigned have been in the Navy for some period of time; it is therefore possible to make use of evaluations of their Navy experience in making assignments. It is of some value merely to know, for example, that a man has graduated from a class A gunners mate school and served as first loader of a 40 mm gun; but it would be considerably more important to know in addition *how well* he could carry out the duties of a gunners mate—how frequently he caused jams through improper loading, how adept he was at

casualty analysis, and how quickly he could replace broken extractors. If he is exceptionally proficient, he might be recommended as gun captain; but if, as is entirely possible, he is shown to be inept, he might be recommended for some quite different type of duty.

Optimal assignments are most likely to be made when adequate measures of *proficiency* are available. Prediction of success in a given activity based on knowledge of past success in that activity is more likely to be correct than is prediction based on aptitude tests or on records which show merely the amount of relevant experience. Advanced classification has an advantage over recruit classification in that the men to be assigned have been in the Navy for some time. It is therefore possible to use evaluations of their Navy experience in making assignments. The first suggestion for improving advanced classification is that greater use be made of measures of proficiency in Naval duties.

Performance Tests

In the case of certain Navy rates, proficiency in the rate can be measured fairly adequately by means of one or two performance tests; often one test is adequate to measure proficiency in that aspect of the duty which is crucial, i.e., which is the most common reason for ineptitude. In some cases these tests are of such a nature that they could be administered at advanced classification centers; in other cases they could most efficiently be given at operational or advanced training activities. The following paragraphs describe several illustrative performance tests.

Radio Code Receiving Tests. A set of radio code receiving tests recorded on phonograph records, which was developed by the Applied Psychology Panel, is suitable for measuring the accuracy with which operators can copy plain language and message-type (coded) material sent at approximately the same speed as is used for U. S. Navy "Fox Schedule" broadcasts.[19] This test would be very useful at a classification center for determining the relative proficiency of radiomen and radiomen strikers and recommending their assignment to particular billets. A certain amount of special equipment (phonograph turntable with pickup and amplifier, headphones, and telegraphic typewriters) is needed, but the test could easily be administered and scored by a Specialist (C). A previous set of tests of a similar nature[5] was used in one Operational Training Command to assist in assigning balanced crews of radiomen to ships.

Signalman Tests. It would be possible to develop tests for signalmen which could be used for similar purposes. Flag hoist spotting tests have already been developed by the Standards and Curriculum Section of the Bureau of Naval Personnel; these consist of colored slides of flag hoists which can be projected, and the men can be required to record the messages. This test and other signaling tests are so far used only in class A signalman schools but probably could be adapted for use at advanced classification centers.

Telephone Talker Test. The only performance test of proficiency which has been used extensively in advanced classification procedures is the Telephone Talker Test,[1] which was prepared by the Applied Psychology Panel. Specially trained men are used in administering this test; the test is given by conducting an interview over sound-powered phones. The testee is required, during the interview, to repeat Naval commands commonly used aboard ship, to read a paragraph containing all the phonetic sounds in American speech, and to speak extemporaneously. He is rated on loudness, rate, articulation, etc.; on listening ability; and on memory span. Finally a rating of "well qualified," "qualified," or "not qualified" is given. The coordination of various departments of a ship through transmitting orders by sound-powered telephone accurately and understandably is obviously of great importance; the elimination of poor telephone talkers from billets which require the use of telephonic communication is essential to the safety of a ship and its success in combat.

Gunnery Proficiency Tests. Some types of proficiency tests require gear and personnel which would make testing impractical as part of the activities of the classification center itself. This need not, however, prevent the administration and use of such measures. The tests

can be given at the termination of the training period in an operational or class A school, and, if the results are entered in the men's service records, they can be employed just as successfully as if given at the classification center by classification personnel.

For example, tests have been developed for measuring proficiency in each of the three jobs involved in operation of the 20 mm gun—the jobs of gunner, loader, and sightsetter.[b] The battery consists of four tests—Dry Tracking, Tracer Position Estimation, Range Estimation, and Dry Loading. Men in destroyer pools are sent to antiaircraft training centers for training in operational gunnery. It would be quite feasible to give the 20 mm gunnery proficiency test battery at the termination of this training and to send the results to the classification center which is attached to the destroyer pool. The classification center could then recommend assignment to gunnery billets on the basis of measured proficiency, rather than merely on such variables as Basic Battery test scores, height, weight, vision, civilian experience, and interests. It is quite feasible to develop and use in a similar manner proficiency tests in operational gunnery for the larger guns for director personnel.

Achievement Testing at a Landing Craft School. A situation which is particularly apt for illustrating the possibility of improving classification by use of proficiency measures was found at a school for training crews of landing craft.[18] At this school the crews are trained for Landing Craft, Vehicle-Personnel (LCVP's). The crew of an LCVP consists of a coxswain, a signalman, a deckhand, and an engineer. (The engineer's billet requires specialized training, not given at this school; so the selection of engineers will not be considered here.) All the men (except engineers) are given essentially the same training for the first four weeks, since it is desirable that any man be able to take over the duty of any other, in event of a casualty. After the first four weeks, training becomes more specialized, with men designated as coxswains spending relatively more time in boat operation, signalmen in learning semaphore, blinker, etc., and deckhands in learning seamanship.

The classification procedure in use was to select men for each billet prior to any training; classification was based on age, physique, vision, previous experience, and Basic Battery test scores. This procedure was used in spite of the fact that the situation was made-to-order for selection on the basis of achievement during the first four weeks of training. The recommendation was made by the Panel project, after certain achievement tests were introduced, that signalmen be selected *after* the preliminary training which is common to all men, and that the selection be based in large part on signaling proficiency as shown by achievement test scores. A similar procedure could be used for selecting coxswains, if suitable measures of proficiency in maneuvering an LCVP after four weeks of training could be developed. The men not chosen for either coxswain or signalman billets would become deckhands, which is the least specialized of the LCVP assignments.

Achievement Testing in Other Schools. The use of grades in class A or P schools and in advanced schools as a factor in selection is a possibility which has not been seriously considered in advanced classification procedures. Unless the school grades are more carefully determined than is usually the case, it is entirely proper not to consider them. However, steps have been taken by Panel projects, working in cooperation with Naval officers, to improve Service school grades, particularly by supplying tests, both of the performance and paper-pencil variety. Given school grades whose meaning is standardized and which are based on test situations resembling the shipboard working situations, it would seem that classification might be improved through consideration of school grade.

Studies of the validity of classification procedures by correlating the measures used in classification with some measures of success in shipboard duties have rarely been made, mainly because of the difficulty of achieving a suitable criterion of performance in the fleet. One study,[16] however, showed that grades made

---

[b] These tests were developed by Project N-106 of the Applied Psychology Panel; the work is being continued by the College Entrance Examination Board under a contract with the Navy Department.

[redacted]

in submarine training courses are of some value in predicting ratings by submarine officers of the performance of enlisted men aboard ship—of greater value, in fact, than any of the other predictive factors studied. This was in spite of the fact that standardized tests of proficiency were not used in determining school grades.

In a number of the Navy's class A schools a definite attempt has been made to improve school grades—by issuing directives describing in detail what is to be included in school grades and, more important, by devising tests by which knowledge and skills can be accurately evaluated.[2,3,9,12,14,17,19] Also in the operational schools of the Amphibious Training Command, Atlantic Fleet, definite progress has been made in developing school grades of increased reliability and validity.[11] The use of school grades which are based on well-designed performance tests and written examinations should be of considerable value in making assignments; provision should be made for standardizing grading systems, reporting the grades to advanced classification centers, and using them along with other measures in making assignments.

If certain of the performance tests on specific pieces of gear (such as disassembly-assembly of a torpedo main engine) are of sufficiently high reliability, it might be well to report their scores separately. The more nearly the measures of proficiency can be made to correlate with specific shipboard assignments, the greater their value to classification centers.

Ratings of Shipboard Proficiency

The measures of proficiency so far discussed have been of the sort that could most efficiently be administered in connection with a training program. Many of the men to be processed in an advanced classification program come directly from the fleet, and considerable time may have elapsed since the termination of training and hence the determination of proficiency by means of performance tests. It is highly important that proficiency measures used for making assignments be *recent;* in the time spent aboard ship following the completion of training, skills may change markedly in either direction. Shipboard experience may result in significant improvement; or, if the shipboard experience was of a type which did not furnish opportunity for practice, the skill may definitely deteriorate. In the case of a radioman, for example, proficiency in receiving code may drop significantly even in the time intervening between completion of training and assignment to a billet, unless provision is made for refresher drill.[22]

The use of performance tests aboard ship is of course possible, and for some billets performance tests perhaps should be given routinely and the results recorded on the man's service record, with the date, for use in later classification. In wartime, however, elaborate routines of test administration aboard ship may not be feasible, and a more convenient procedure for evaluating performance is needed. The use of rating methods would provide the means for making such evaluations.

The Need for Training of Raters. The common errors in ratings (halo effect, error of leniency, etc.) are well enough known that a detailed discussion of them is not necessary here. It is probably sufficient to say that ratings of proficiency in shipboard jobs will be useful in advanced classification only to the extent that care is taken in planning the rating scales and training the raters in their use. Only through careful work can ratings of acceptable reliability and validity be obtained.

The training of officers or petty officers who make the ratings is essential in order to obtain valid and reliable ratings. Training should include discussions of errors such as halo effect, overrating, and prejudice, as well as an adequate description of the distribution of abilities and of the scale itself. That the results of using rating scales without opportunity to train raters or to review and follow up the raters' evaluations is likely to be of limited value is shown by one recent study;[16] a factor analysis study of the rating scale traits revealed that the ratings were characterized by a great deal of halo, and that almost any one of the traits could be used to represent the whole of the ratings. This was in spite of the fact that the traits rated were apparently quite heterogeneous, including such traits as adaptability, courage, sociability, and dependability.

[redacted]

In order to avoid such results as were obtained in the study referred to above, one should not only make adequate provision for training raters, but also one should take considerable care in planning the rating scale itself. In general, the traits to be rated should be as objective and well defined as possible. Probably a graphic rating scale, on which certain points are defined with descriptive terms, is in general most satisfactory. Each trait should refer to a single type of activity. The effort should always be made to have each man rated by a *number* of his supervisors or officers, so that the final score can be based on the consensus of opinion and not on one person's judgment.

Work Readiness Tests

Even with a fairly elaborate procedure for entering performance test scores and ratings in the service records of enlisted men, there are likely to be cases where data needed by a classification interviewer are lacking or out of date. He may be unable to make an assignment without resorting either to records of aptitude test scores and the like or to trusting statements made by the interviewee regarding his own training and experience. Tests similar to the oral trade tests of the U. S. Employment Service, but based on Navy jobs, would be particularly valuable in such situations.

The Applied Psychology Panel has made a beginning in the development of tests of this type. One Panel project has developed such a test (termed a work readiness test) for distilling plant operators.[13] This test consists of 15 items which may be administered orally or as a written test. Two forms were prepared which are equivalent with respect to difficulty, content, and discriminating power. The test has satisfactory reliability. It differentiates men with watchstanding experience on distilling plants from men with nonengineering rates and also from men with engineering rates but no distilling plant experience. Apparently neither chance nor experience other than distilling plant operation is likely to produce spuriously high scores. Use of the test thus furnishes a method of evaluating the claims of those who profess to be experienced in distilling plant operation. The test is objective and economical of time, and the results can easily be interpreted in terms of assignment to training or to shipboard duty.

The availability of a battery of work readiness tests covering a variety of billets would be of considerable value to a classification interviewer; such tests could be administered in cases where more direct measures of proficiency are not at hand. The procedures for developing work readiness tests are described in an Applied Psychology Panel publication.[15]

13.3.2 Extension of Advanced Classification Procedures

In order for advanced classification to be of maximum effectiveness, it should involve more than the making of recommendations, which may or may not be followed, only at relatively infrequent stages in the career of the enlisted man. There should be throughout the Navy sufficient appreciation of the aims and methods of classification at least for sympathetic cooperation with advanced classification. It would be still more desirable to have officer personnel at all ships and stations who are qualified through interest and training to undertake classification procedures as an extra duty if not as a full-time assignment. In addition, all other officers, who do not serve as classification officers in a direct capacity, should nevertheless know enough about classification to use sensibly the data in a service record for assigning or advising enlisted men, and to realize the importance of keeping service records complete and up to date.

One of the Applied Psychology Panel projects made a survey, in connection with a study on the validation of selection procedures, of the extent to which classification data were available to ships' officers on destroyers in shakedown.[8] It was discovered that Division officers charged with the responsibility of assigning men in most cases were not aware of the existence of the Q-card, and of those who did know of its existence few were aware that it contained information which could be used in making assignments. As long as this situation remains typical, it means that classification

data are of extremely limited value. Not all problems of assignment, training, transfer, and the like which occur aboard ship can be handled through recommendations made at a shore-based classification center.

The study also revealed that Q-cards were frequently not available, that test records on Q-cards which were available were frequently missing or incomplete, and that Q-cards were often found for men not aboard. In other words, considerable laxity in connection with classification procedures was evident.

Steps have been taken by the Enlisted Classification Section which, if followed up effectively, should lead to considerable improvement in the attainment of the objectives of advanced classification. The provision for training one officer and one yeoman from each ship (see Section 13.2.3) is particularly important in this regard.

It might be desirable in addition to have classification officers attached to Precommissioning Training Centers whose duty is to serve as temporary "technical aides" to the executive officer of each new ship. Such an officer could serve on each new ship from the time the crew was processed until the ship reported for shakedown. During this time he could work with the officer designated as classification officer for the ship, assisting in problems of classification as they arise and incidentally providing indoctrination for the officers and yeomen. This plan has the advantage of providing instruction at a time when it is especially needed and consequently at a time when motivation to learn should be high.

For long-term effectiveness, the indoctrination of officers in general can best be achieved by the inclusion of definite courses in personnel procedures at the U. S. Naval Academy. This would eventually ensure some understanding of the problems of classification among high ranking officers in the Navy, as well as among engineering, deck, gunnery, and other types of officers throughout the Service. Only when the specialized training of classification officers is supplemented by general understanding and cooperation among other officers can the important goals of personnel classification be achieved. Only then can classification become a *continuing* process through which the service record becomes more and more valid as a device for making assignments and enlisted men continue to be given training and billet assignments which make them of maximum value to the Navy.

As has already been mentioned, recommendations from classification centers are occasionally ignored, and billets may be assigned by the personnel department on the basis of the alphabetization of names, for example. It is not uncommon for assignments aboard ship to be made by assigning men to those battle stations which are nearest their bunks. In such situations a classification officer can do nothing, since billet assignments are actually a function of the personnel department. Similarly, classification officers have no jurisdiction over the crew of a ship; once a man has been assigned to a ship, the commanding officer has jurisdiction over his future assignments. As a result, it sometimes happens that the poorest men tend to be assigned to advanced training, since the commanding officer is loath to give up a good man. One ship may have more than enough good first class radiomen while another ship has none. When a request for transfer is received, the commanding officer may circumvent the purpose of the transfer by getting rid of the poorest of his first class radiomen. In defense of this situation it is often said that the officer in charge of his ship depends upon his crew for the safety of his ship and its success in combat. While no one would gainsay this position, the long-term values to the Navy of equitable distribution of men among its fighting ships and of selecting the best qualified men for training should not be overlooked.

Considerations such as these might lead to the suggestion that the responsibility for assignments within a particular region (such as a Naval district or a fleet command) be placed in the hands of a well-qualified classification officer. He would have authority to see that classification recommendations are followed and that men are equitably distributed among the units of the fleet, and to make sure that all directives and policies regarding classification are followed. By making use of methods of evaluating proficiency, the classification officer

[redacted]

can objectively control the equitable distribution and proper assignment of the men under his jurisdiction. Through such a method, which would neither be too centralized nor give too much autonomy to separate ships and stations, the classification procedures might be made to work to the greatest advantage to the Navy as a whole.

13.3.3 Procedures for Coding and Filing Personnel Data

In the days before radio, radar, sonar, loran, fire control equipment, and automatic weapons, naval warfare was a much simpler matter than it is today. The use of Navy ratings at such a time was probably a fairly adequate device for classifying and selecting personnel. With the increased specialization that has come with a highly mechanized Navy, however, the value of rates for selection is highly questionable. A first class gunners mate, for example, may be entirely capable of operating and maintaining .50-caliber Browning machine guns but be relatively ignorant of larger automatic weapons such as a 40 mm or 5″/38 gun; he may have never served on a ship which mounted the larger weapons. Similarly an electrician's mate may be quite capable of ringing out and repairing lighting circuits but be ignorant of selsyn units and interior communications.

One possibility is to increase the number of ratings and to add designators to ratings which indicate fields of specialization. This method has been used to a certain extent. New Navy rates, such as sonarman and radarman, have been added, and the designators (M) and (R) have been added to fire controlman (FC) to indicate whether the man is qualified for maintenance or for operation of fire control equipment. This is at best only a partial solution, however, unless the system is extended considerably. For example, in assigning a fire controlman (M) to a particular billet, it might be necessary to know his qualifications for maintaining Mark 37, Mark 51, and Mark 52 gun directors, rangefinder optical equipment, torpedo directors, fuze setter and sightsetter equipment, and other types of computers, stable elements, and the like. The problem is just as complex for many other rates. A machinist's mate may have specialized in high or low pressure steam turbines, reciprocating steam engines, diesel or gasoline engines, or turbo-electric equipment. A radarman may be qualified on only one of a number of different types of radar sets.

In other words, it would seem that what is needed is a method of *coding* Navy occupational skills which can be used efficiently in classification procedures. A five- or six-digit code would probably suffice, in which various digits or pairs of digits could be used to represent rate, pay grade, field or fields of specialization, and degree of proficiency in the field of specialization. Classification would be facilitated by appropriate use of such codes both at shore-based stations and on ships. At shore stations the code could be punched into electric accounting punched cards, and sorting of the cards representing the men in a pool could be quickly made. This would rapidly make accessible the data needed for making billet assignments and would require much less individual study of Q-cards. For shipboard use,[10] the code would permit the use of such a device as the McBee Keysort system.[6] The Classification and Selection Section of the Bureau of Navy Personnel[23, 24] has developed a Manual of Navy Enlisted Job Classifications which provides a Navy occupational skills code. This code went into effect July 1, 1946.

[redacted]

Chapter 14

CONSTRUCTION AND STANDARDIZATION OF GROUP TESTS

By *Norman Frederiksen*[a]

Summary

THIS CHAPTER consists of an outline of the procedures recommended for constructing aptitude tests for the selection of military personnel.

14.1 INTRODUCTION

The procedures which the Applied Psychology Panel used and recommends for use in preparing tests for the classification of military personnel are described in this chapter. They apply particularly to the development of aptitude tests of the paper-and-pencil variety which are power tests rather than speed tests. The procedures are described here in detail because they are not generally available. They are described with refinements which could not always be used because of the speed demanded by wartime pressure. The best tests constructed under Applied Psychology Panel direction, however, were those in which the recommended procedures were followed most closely; when short-cuts were used, the tests suffered by comparison. In peacetime test construction, thoroughness in the application of the methods described is strongly recommended.

It is impossible to construct valid tests merely by sitting at a desk and writing test items. Before any work on test preparation is done, it is necessary to find out what the proposed test is to be used for—what it is supposed to predict. It is necessary to study the to-be-predicted activity in order to determine the critical skills and abilities required for success. It is further necessary to acquire at least a general notion of the characteristics of the group to which the test is to be given, and to become acquainted with the conditions and limitations of the classification program of which the test is to become a part.

[a] This chapter is based on the work of NDRC Project N-106 and the procedures followed by recognized test experts.

Only after these things have been done can one start writing test items. The items themselves must then be tried out and an item analysis made. It is necessary to select items for the final test in terms of their difficulty and validity, to determine time limits, to make certain that instructions are clear, and to arrange the items in proper order. All these things are necessary before one can say with confidence, "Here is a good test."

14.2 PRELIMINARY DETERMINATION OF WHAT THE TEST IS TO ACCOMPLISH

14.2.1 The Performance to Be Predicted

The first task is to decide specifically what performance is to be predicted. This performance may be something as specific as using a stereoscopic heightfinder, or something as general as the intellectual ability to pass any Service school course.

DETERMINING WHAT IS TO BE PREDICTED

The most general problem likely to arise is whether a classification test battery should be aimed at predicting Service school success (the immediate goal of trainees) or eventual success in carrying out a military job. This problem arose in connection with the preparation of the Navy's Basic Classification Test Battery. The decision made then was to attempt primarily to predict Service school success, on the grounds that it was especially important to prevent waste of time and equipment in trying to train men who were poorly qualified with respect to this first goal. At the time the decision was made, the choice may have made some difference in the types of tests which were developed; but with improved, more realistic training in Service schools, such that the activities in Service schools resemble more closely the duties

aboard ship, it would make less difference which criterion one aimed to predict.

Determining the Test Content

After deciding what is to be predicted, it is necessary to select the kind of material to include in the test. The second step is to make a thorough job analysis of the performance to be predicted. (See Applied Psychology Panel, Volume 2, Chapter 14.) In some cases the results of a job analysis will already be available. These analyses should be studied and used, but acquiring firsthand information about the job should not be neglected. If a test is to be made to predict success in a signalman school, for example, the test specialist should visit such schools, study the curricula, analyze the causes for failure, and if possible actually take the course himself to find out at first hand what the problems are like. Only by such firsthand experiences can he really get the feel of the situation and the background which best equips him to form hunches as to the types of items most likely to be successful.

Evaluation of Training. At this point it may be mentioned that the test specialist has an opportunity and a responsibility to make evaluations which have implications beyond the field of classification. He may find that training is not realistic and that tests constructed to predict school success will actually select men poor in performance of their military duty. For example, one school curriculum led to the selection of bookish men, while a quite different type was most successful on the job. (See Chapter 17 of Volume 2, Applied Psychology Panel.) Obviously in such a case it would be extremely shortsighted to go ahead and prepare a test which would predict only school success. Instead steps should be taken to make the training more realistic, so that both training and selection are better oriented toward the ultimate goal of maximum quality of performance in combat.

The Availability of a Criterion

Before beginning the work of test construction, plans should be made for validating the test. Validation requires that a reliable and meaningful criterion of success on the job be available. In many cases it will be found that a suitable criterion is lacking or that the available criterion (such as passing or failing a Service school course) is itself probably not a valid indication of success.

If a suitable criterion is lacking, it will eventually be necessary to develop one. If the development of the criterion measure is not begun until after the predictive test has been completed, the validation of the test will be considerably delayed. Serious consideration should therefore be given to the possibility of beginning work on the development of a criterion measure at the same time that the construction of the test is started.

14.2.2 Analysis of the Classification Situation

The Testing Situation

In addition to studying the performance in which success is to be predicted, it is also desirable to investigate the situation in which the proposed tests are to be used. Factors such as time available for test administration and test scoring, size of groups to be tested, use of interviews or other classification procedures, and physical facilities for conducting a testing program have a significance with respect to the nature of the tests to be developed.

Considerations such as those suggested above should not be allowed to have too much influence in determining policies with regard to test construction. During peacetime developmental work, it should be possible to relegate such factors to a secondary role and to shape the situation in which the tests are to be used to fit whatever tests are found to be useful. On the other hand, one should guard against setting up a testing organization which, however suitable it may be for peacetime needs, is too elaborate and time-consuming to permit classification of recruits at a rate which might become necessary during mobilization for war.

Characteristics of the Group to Be Tested

Advance information concerning the characteristics of the group to whom the test is to be given will be of value in designing tests and

[redacted]

will help in gauging the proper difficulty of test items. Knowing the central tendency and variability of the group with respect to age, education, vocational training, and occupational experience should be of value in this connection. Score distributions and item analyses of tests previously taken by the group will be particularly useful if they are available.

14.3 PREPARATION OF THE FIRST EXPERIMENTAL TESTS

The first version of a test should be thought of as purely experimental; it should be designed to give as much information as possible to aid in future revisions of the test. This section includes specific suggestions with regard to planning the first experimental version of a test for maximum usefulness in obtaining data for subsequent revision.

14.3.1 Writing the Test Items

On the basis of the preliminary survey of the performance to be predicted, the psychologist can prepare the items included in the first experimental test. The first version is in the nature of an experiment, and all reasonable hunches should be followed up. This means that not one, but several tests or subtests should be constructed, with the idea of following up the best leads suggested by the preliminary research findings on comparative validity, reliability, and internal consistency. The findings reported in the literature by previous investigators should be studied for ideas with respect to suitable test items.

Now for some more specific suggestions with respect to the preparation of the first experimental tests:

1. Make three to five times as many items as will be used in one test. A sufficient number of items should then prove satisfactory to permit the selection of matched items for at least two parallel forms of the test. For some kinds of items (such as arithmetical computation) where comparatively few of the items will fail to meet the standards for selection, fewer items need to be prepared than for types of tests where successful items are more difficult to write, such as verbal analogies and mechanical comprehension. The number of items that can be included in one test is limited, of course, by such factors as fatigue and boredom of the testees.

2. Edit the items carefully. A lower mortality rate will result. Too often it is found that an item which must be rejected because of item analysis results contains a defect which could have been discovered in advance. The editing should include not only careful study by technical experts (such as the study of mechanical comprehension items by engineers) but also careful inspection from the point of view of rhetoric and ambiguity.

3. Strive for a fairly wide range of difficulty in order that items can be selected which have a suitable range of item difficulty. Information on the general ability level of the group to be tested will help in this connection—especially information on tests of similar nature which have been administered to that group. Experience has shown that it is harder to construct valid *difficult* items than valid *easy* items;[22] therefore a greater excess of difficult items than of easy items should be prepared, in order to have available a sufficient number of valid difficult items.

4. Choose the type of item—multiple-choice, matching, true-false, completion, or other type—which appears to be most advantageous in the particular situation. In making the choice, consider such factors as time and space per item, ease and objectivity of scoring, freedom from chance success, and ease of giving directions, as well as the appropriateness of the type of item for the kind of subject matter being dealt with.

All things considered, the multiple-choice type is in general most satisfactory. Tests composed of multiple-choice items are easy to administer, they are adaptable to machine scoring or are easily scored by hand, and with four or more choices are sufficiently free from the factor of chance. A large number of such items can ordinarily be administered in a short time. Most important, the multiple-choice item is adaptable to a wide variety of purposes, from

[redacted]

measuring simple recognition to complex thinking. However, for certain kinds of subject matter other types of items may be superior. That type of item should be chosen which is best adapted to the type of measurement being attempted.

5. For each item, prepare two or three more distractors (wrong answers) than are planned for use in the final form of the test. It will then become possible to select those distractors which function best.

6. Base the selection of distractors on objective empirical evidence when feasible. In order to do this, the psychologist should prepare a free-answer form of the test (i.e., a completion or short-answer test in which each testee writes his answer to each item). The free-answer test should then be administered to a group similar to the population eventually to be measured. The frequency with which various answers are given can be tabulated and the most frequent wrong answers selected as potential distractors for the multiple-choice form of the test. By use of a more elaborate item analysis of the free-answer test, it is possible to select items and distractors on the basis of their discriminative value. It is further possible to estimate the difficulty of each item on the basis of the frequency of correct answers.

This method is obviously not applicable to all types of tests; it could not, for example, be used to good advantage on a test where the choices are represented by pictures. The method is applicable when there is the greatest similarity between the operations performed in a free-answer and a multiple-choice form of the test. Arithmetic or mathematics tests are especially adapted to this technique of obtaining distractors.[21]

7. Try to present the test material as realistically as possible; for example, the material in arithmetical reasoning items should be consistent with the experience of the testees to as great an extent as possible. Use pictures whenever they are appropriate. When testing a group with a wide range of ability, it is difficult to avoid high intercorrelations among tests. By avoiding verbal items except where they are especially appropriate, the intercorrelations may be reduced.

14.3.2 Planning the Format and Directions

A group test cannot be successful unless the directions for taking it can be followed. Many testees will have had little previous experience in taking examinations and will have a rather low level of ability in following either oral or written directions. It is important that a man's test score represent his ability with respect to what is being measured by the test and not ability to follow complicated directions on coding correct answers or using separate answer sheets. The format of the test and the oral and written directions should therefore be as simple as possible, and format and directions should be studied during the use of the experimental versions with their later improvement in mind.

Format of the Test

All details should be carefully studied from the point of view of ease of understanding and following the test procedures. The choices for an item should be arranged in a column instead of being strung out along the page, even though more space is required. This arrangement makes clearer what identifying letter or number goes with each choice. If the items are numbered, the choices should be lettered, so that in recording answers there is a minimum of confusion and error; (item) *4*, (choice) *C* is easier to bear in mind while recording on a separate answer sheet than *4, 3*, which might easily be recorded as *3, 4*.

Use of Separate Answer Sheet

Probably it will be decided that the test should be scorable by means of a test scoring machine; if so, the test blanks for the preliminary forms of the test should have answer sheets which are good facsimiles of, if not actual, machine scorable answer sheets. The machine scoring feature does not necessarily involve the use of a separate answer sheet; in some cases, it is feasible to print the test itself on the answer sheet.[21] This would eliminate the necessity of coding the choices and making answers according to this code, which is undoubtedly a source of some error. If the entire test can be printed on the answer sheet, this

method definitely should be employed. If several pages are required, however, it becomes necessary to add scores from different pages, which introduces the likelihood of large errors; consequently the separate answer sheet would be preferable in such a case.

Writing the Directions

The writing of clear yet concise directions requires considerable care and insight into the thinking of the none-too-bright recruit who has perhaps never before taken a standardized test. The directions should be written in such a way that the test procedures can be correctly followed by all recruits—the dull as well as the bright. Test scores should not reflect differences in ability to understand directions.

Perhaps the best approach in writing test directions is to describe, one at a time, the specific procedures which must be followed in the order in which they would most naturally occur. The recruit can be carried through the procedures most effectively by requiring him to participate actively through answering and recording answers to several practice questions. Sentences should be short, and vocabulary should be at a level which is low enough for all but the most illiterate to understand. In most cases, and especially for tests which necessitate rather complicated procedures (such as Surface Development), it will be of considerable help to pretest the directions, using a sample group which includes many low-ability recruits.

The directions should be very explicit in requiring the testee to answer every question. If this is not done, it will be necessary to correct the scores for differences due to chance success. Even then, final scores will depend in part upon the men's willingness or unwillingness to guess. The men are more likely to follow a common policy if guessing is encouraged than if it is discouraged. The word *guess* may be undesirable, but the same result can be achieved by saying, "If you don't know the answer, mark the one you think is most likely to be right."

Making the Experimental Tests of Maximum Value

A number of precautions should be observed in planning the first experimental tests to make the item analysis data of maximum value. In the first place, the format of the test and answer sheet should be such that part scores are readily obtainable, so that items of different types can be studied separately. Items of different types can later be thrown together in an omnibus form, for example, if desired; but it is important to be able to study separately the efficiency of items of various types. Furthermore, it is important to attempt to plan the test in such a way that data on the various items are equally trustworthy, i.e., affected as little as possible by such factors as fatigue, monotony, rate of work, or attitude. One way to minimize the effects of such factors is to publish the experimental tests in two or more forms which differ in the order of the items. Thus items which are near the end of the test in one case will be near the beginning in another.

Ideally, the time limit should be so liberal that at least 90 per cent of the recruits are allowed to finish; otherwise the data available for the terminal items are likely to be based on a select group. Unless time limits can be determined by pretest experiments, it may be desirable to leave the time limit undecided until the test is actually given. Then the test can be terminated when the desired proportion has finished. The number still working at any moment can be determined by requiring the men to indicate in some way, such as by putting the test paper on the floor, when they have finished.

14.4 ADMINISTRATION AND ANALYSIS OF THE FIRST EXPERIMENTAL TESTS

14.4.1 Administration of the Test

After the test booklets and answer sheets for the experimental test have been prepared, the test is ready for its trial administration. The sample to which it is given should resemble as much as possible the population with which the test is ultimately to be used; otherwise estimates of item difficulty (in particular) may be misleading.[18] In order that the various statistics used to evaluate test items may have satisfactory reliability, the sample should be made up of a minimum of 500 men. If different

[redacted]

forms of the experimental test (containing the items in different order) are used, they should be given to subgroups of equal size selected at random from the sample. The tests should be well proctored, and the proctors should be instructed to notice particularly behavior of the testees which indicates how successful the test directions have been. Questions by the recruits would be such an indication of the adequacy of the directions. A record should be kept of any such evidence which is forthcoming. The test should not be terminated, if possible, until at least 90 per cent of the men have completed the test.

14.4.2 Item Analysis of the Test

The test is now ready for tabulation of responses and item analysis.[19] The analysis of the test should furnish for each item some measure of its difficulty and some measure of the validity of the item (i.e., its relation to some criterion, usually the score on the test or subtest of which the item is a part). In addition, the item analysis should show how many individuals chose each alternative answer to each item and the average score (on the criterion) of the group choosing each alternative. Finally, the reliability of the test should be determined by correlating random (usually odd-even) halves of the test and correcting for length by means of the Spearman-Brown prophecy formula. Once these computations have been completed, the first revision of the test can be made.

The Analysis Procedure

The amount of item analysis information available about each item and about each alternative within an item determines the adequacy of the revisions that can be made. A very detailed item analysis procedure[19] was, therefore, followed in constructing the Navy tests described in Chapters 2 and 3.[b]

Figure 1 shows a sample item analysis sheet. It gives the information which was regularly used in making final choice of which items to include and which to reject. The types of information made available and their interpretations are:

Item. The item itself is typed on the sheet. For ease in sorting, its number is also printed in the box at the upper left.

Distribution of Responses. The number of men selecting each alternative and the number omitting the item entirely are shown in the column headed "N." Three men omitted item 96. Two hundred and eighty-eight answered it correctly by choosing the first alternative. The heavy lines above and below row 1 indicate that that was the correct answer. The other 209 men failed the item. Rows 2, 3, and 4 supply information on how many chose each wrong alternative.

Number of Cases. The *Base N* is given at the top. The number who actually attempted an item (N_t) appears in the bottom row of the table. Anyone who has answered a particular item, and anyone who has answered a subsequent item, is considered to have tried that item. The sample item shown in Figure 1 was placed early enough in the test so that N_t equaled Base N. In items placed near the end of a test, N_t is frequently less than Base N. N_t rather than Base N is used in computing the statistics described in the following paragraphs.

Item Difficulty. Two measures of item difficulty are used. The simpler is symbolized by p. It consists of the percentage of N_t which answered the item correctly. In this case $p = .58$. The second difficulty measure is symbolized by the Greek letter delta (Δ). Δ is expressed in terms of transformed criterion scores. The essential features of these transformed scores are that they correlate 1.00 with the original scores, that the mean of the total sample on the transformed scores is uniformly 13.00, and that the standard deviation of the total sample is uniformly 4.0. Δ is defined as that transformed criterion score above which the percentage of cases equals p. The more difficult the item, the higher the value of Δ. In this sample $\Delta = 12.2$ and $p = .58$. For an item with a Δ value of 13.0, p would equal .50.

Correlation with Criterion Score. The biserial correlation between the item and the criterion is computed. In this case it equaled .60.

[b] Reference 19, *Characteristics and Uses of Item-Analysis Data*, has been declassified and will be published by the Stanford University Press in the series of Applied Psychology Monographs.

[redacted]

ITEM ANALYSIS

Card Number 48

ITEM 0 9 6

TEST MECHKN

FORM I

BASE N 500

COLLEGE ENTRANCE EXAMINATION BOARD
Research and Statistical Laboratory
Princeton, New Jersey

Date Tabulated 6 44 Operator Number 9

Response	Code	n	Σx	Mean	Σx²
	0	3	16		88
	1	288	4215	14.635	65683
	2	71	848	11.9	10814
	3	45	458	10.2	4990
	4	93	965	10.4	10995
TOTAL TRIED (t)		500	6502	13.004	92570

$\sigma_t =$ 4.004

$p =$ .58

$M_+ - M_t =$

$\dfrac{M_+ - M_t}{\sigma_t} =$

$\dfrac{p}{z} =$

$r =$ $\boxed{.60}$

$\triangle =$ $\boxed{12.2}$

$p = \dfrac{n_+}{n_t}$; $\sigma = \sqrt{\dfrac{\Sigma x^2}{n} - M^2}$

$r = \left(\dfrac{M_+ - M_t}{\sigma_t}\right)\left(\dfrac{p}{z}\right)$

96. In a gasoline engine, the gas mixture should explode in -- (a) the cylinder (b) the intake manifold. (c) the exhaust manifold. (d) the carburetor.

Computed by K

Checked by

NOTES:—

1. x', based on p, is the distance from the mean along the baseline of the normal curve, in terms of unit standard deviation. If p is less than .5, x' is positive; if p is more than .5, x' is negative.
2. Compute means only for responses made by ten or more candidates.
3. Compute M_+ and M_t to three decimal places; means of all other responses, to one.
4. Record r to two places; $\triangle$ to one.
5. Carry all other computations to three decimal places; i.e., $\dfrac{\Sigma x^2}{n}$, $\dfrac{\Sigma x^2}{n} - M^2$, σ_t, p.

FIGURE 1. Sample item analysis sheet.

[redacted]

The criterion employed is the score on the subtest of which the item is a part; if the test is not divided into subtests, then the score on the total test is employed.

Information on Alternatives within Each Item. The mean transformed score of the men selecting each alternative is given in the column headed "mean." The mean score of the 288 men who selected the correct alternative was 14.635, substantially higher than the mean of the entire group (13.0). The mean transformed scores of the men who chose the wrong alternatives were all well below 13.0. Alternatives *c* and *d* (mean scores 10.2 and 10.4) appealed to a poorer group than alternative *b* (mean score 11.9).

Alternative Criteria for Item Analysis

If it is possible to administer the trial test to a group which satisfies the criteria for a representative sample and at the same time can be used as a validation group, not only can the test be validated more quickly, but also a supplementary item analysis using the external criterion can be employed; this permits the selection of items on a dual criterion: internal consistency and correlation with the external criterion (see Chapter 15).

Still another type of analysis would be of value in situations where, as in developing a battery of tests, it is desirable to reduce the intercorrelations of the tests as much as possible. This cross item analysis consists of correlating the items of each test with scores on all the other tests in the battery, thus permitting the selection of items on the basis of high correlations with their own test and low correlations with other tests. Such a procedure would be particularly useful for recruit tests, where the range of talent is great, and for tests where low correlations are desirable but difficult to obtain, as in tests of verbal facility and arithmetical ability.

Selection of Items for Alternative Forms of a Test

It is always desirable to have two or more alternative forms of any test which is to be used widely and repeatedly. The construction of alternative test forms always poses a number of statistical problems. The distributions of scores on the two forms should be similar. The difficulties of the items should be similar. Item-test correlations, reliabilities, and other statistical characteristics should be matched for the two forms of the test. The satisfaction of these requirements requires careful planning and considerable work.

The procedures to follow in developing equivalent forms of a test are in general those described in this chapter. References 17 and 18 describe in detail the application of these procedures to the special problems of developing equivalent test forms.

Uses of Item Analysis Data

The principle uses of item analysis data may be briefly summarized as follows:[19]

1. Item analysis supplies detailed, objective, quantitative information for each item. This information cannot be obtained by "expert judgment" nor by any manipulation of the reliability coefficient.

2. The objective, quantitative information from item analysis is well suited to help settle arguments or objections concerning specific items; and it provides a convenient, practical basis for selecting items for subsequent forms of a test.

3. Item analysis data provide information which is useful in revising and improving test items.

4. The distribution of item difficulty can be improved with respect to symmetry, continuity, and average level, on the basis of the evidence provided by item analysis concerning the difficulty of each item.

5. The reliability of the test may frequently be improved by the judicious selection of items on the basis of item analysis data.

6. The independence of the test from other tests in the battery may be improved by the application of a cross-item analysis technique.

7. The external validity of the test can frequently be improved, if the item analysis includes the correlation between each item and a valid external criterion.

8. Item analysis data stimulate hypotheses and insights useful in construction of tests and interpretation of test results.

[redacted]

14.5 REVISING THE TEST

The revision of the test is a matter of selecting items on the basis of the item analysis statistics and modifying the format and directions.

14.5.1 Selecting Test Items

Determining the Number of Items Needed

On the basis of the reliability coefficient which was obtained for the trial test, it is possible to determine about how many items should be included in the test. Ordinarily a reliability of about .90 is desirable; the number of items necessary to secure this reliability can easily be estimated by means of the Spearman-Brown formula. Various other considerations may of course lead one to revise that number in one direction or another. It might, for example, be desirable to decrease the number of items somewhat in order to avoid excessive time for test administration or to avoid using more than one side of an answer sheet. It should be remembered that the estimate of the reliability of the test will probably be an underestimate, since the selection of test items should ordinarily result in a test which is improved with respect to internal consistency and reliability.

Criteria for Item Selection

Items are then selected on the basis of a multiple criterion which includes the difficulty of the items and their relationship to the total-test score. The most desirable items are those which have the highest correlations with the total-test score and whose difficulty values are scattered near a level halfway between a chance level and the level where all testees answer the item correctly[23] (e.g., for five-choice items, those which are passed by about 60 per cent of the group). The most desirable scatter about this level of difficulty is not entirely clear as yet; but from practical considerations it would seem desirable to include a few relatively easy items as "icebreakers" at the beginning of the test and to avoid an excessive number of items which are so difficult as to have an adverse effect upon the attitude of the testees. For a five-choice test, the items should probably range in difficulty from those answered correctly by about 85 per cent to those answered correctly by about 35 per cent, with the majority of the items at about the 60 per cent level.

Assuming a fairly narrow distribution of item difficulties, the optimal values for item–total test correlation is a question not fully answered as yet. Presumably no serious errors will be made if one selects the items with the highest correlation with the total test, provided a sufficient number of items are available whose biserial correlations are above, say, .25. If more than enough items are available which are fairly homogeneous, as indicated by uniformly high biserial correlations, it is possible to select on a judgmental basis, choosing those items which best represent the hunch being tested.

The lower the item–total test correlations tend to be, the less value they have for item selection. If they are all extremely low, they indicate merely that the items are heterogeneous. In such a case items can be selected on the basis of subjective judgment as to similarity. A better technique would be to subject the items to a factor analysis and select items on the basis of their loading with a particular factor. The factor analysis might be based on a subjective classification of the items into groups which appear to be alike; the scores on the groups, treated like separate test scores, could then be correlated and factor-analyzed.

Selection of Distractors

It was also recommended previously that the items of the trial form should contain more distractors than are needed for the final form of the test, in order that the best distractors can be chosen. The recommended item analysis procedure would supply the necessary information, viz., (1) the number of people who choose each distractor and (2) the mean score on the test for each group. The most efficient distractor is one which attracts a fairly *large* group of *low-ability* men; therefore those distractors should be eliminated which relatively few choose, and those which are chosen by high-ability men more frequently than by low-ability men. The relative importance of these two factors is difficult to evaluate in some instances; for example, it might be difficult to

[redacted]

decide whether to eliminate choice *a* or *b* in the following example:

Alternative	Number choosing each alternative	Mean score of those choosing each alternative
a	37	14.1
b	14	13.0
c	31	10.6
d	24	10.4
e	62	12.3
Correct answer: *f*	116	17.3

Alternative *a* is chosen by a considerably larger group than *b*, but those who choose *a* make scores which are on the average considerably higher than the scores of those who choose *b*. (The means are in terms of standard scores based on a distribution with a mean of 13 and a standard deviation of 4.) No rule can be presented to govern a situation of this sort; further research, dealing with different types of test material, is needed.

REVISING ITEMS

In cases where the experimental test fails to yield a sufficiently large supply of items or when additional forms of the test must be prepared, the rejected items should not be overlooked as a source of supply.[19] Since a considerable amount of statistical information concerning the item is available, the chances of finally achieving a satisfactory item through revision are often superior to the likelihood of writing a successful new item.

The data may show, for example, that some distractors are successful—they draw a goodly number of men whose scores are on the average well below the average for those who got the item right. Such distractors should be retained. The failure of the item may be due to a distractor which draws a high-ability group; if so, it may be sufficient to rewrite this distractor. In other cases it may be necessary to rewrite distractors to make them somewhat more attractive, so that they will become functional parts of the item. It is also possible, of course, to revise the part of the item preceding the choices; this would mean that the statistical information would be invalidated. After revision, the items are ready for a second trial through their incorporation in another experimental form of the test.

THE USE OF EXTERNAL CRITERIA IN ITEM SELECTION

The above description of the procedures for selecting test items will perhaps be adequate for the situation in which the test is analyzed using only total-test score as the criterion. With an external criterion, a somewhat more elaborate procedure of item selection must be used.[15] This method is described in Chapter 15. Use of an external criterion is most useful when there is only one specific criterion to be predicted; in cases where success in any school or any of a limited number of schools is to be predicted, the method would be less practicable.

Where the items of a test have been correlated with other aptitude tests as well as with the test of which the item is a part, it is possible to minimize the intercorrelations among the tests of a particular battery, by selecting items which have high correlations with their own test and low correlations with the other tests.

14.5.2 Revising the Directions and Format

REVISING DIRECTIONS

If one has been careful to observe the performance of the recruits with respect to their understanding of the test directions, the modifications which are required should be apparent. In scoring the tests it may be noticed where common errors are made, such as recording answers in the wrong column of the answer sheet or confusing the spaces for practice questions with spaces for the test proper. If evidence of such errors is found, appropriate changes in format or directions should of course be made. The revisions might take the form of modification of written directions, boxing-in the practice questions and answer spaces, or changing the method of coding answers.

DETERMINING TIME LIMIT

It may be possible at this point to determine precisely the time limit for the test, on the basis of observations of rate of work on the

[redacted]

experimental test. More likely, however, it will be necessary to establish the time limit on the basis of further experimental administrations of the revised test to representative samples of the population. In the case of a power test, the time limit should be such that additional time would not affect the test scores of students to any great extent.

Arrangement of Items

It was recommended above that the number of items in the trial test should exceed the number required for one test by a considerable amount, in order that parallel forms of the test could be prepared with a minimum of effort. The two forms of the test should contain items which are matched with respect to content, difficulty, and correlation with total score, and should resemble each other as closely as possible with respect to means of item difficulty values. If sufficient items which satisfy the above criteria are not available, of course only one form of the test can be obtained, and more work will be required in order to try out additional items.

The items which have been selected for each form should be arranged in order of difficulty as determined by the item analysis. This arrangement will permit a somewhat shorter time allowance than would otherwise be necessary, since there is ordinarily a sufficiently high correlation between level of ability and rate of work so that the men who have no opportunity to attempt the last items would be unable to answer them correctly anyhow. Having the items in order of difficulty has the advantages of permitting the men to build up confidence at the beginning and enabling them to attempt the majority of items that they have the ability to answer. With items in a different order, the men will waste time trying to answer items which are too difficult for them while leaving easy items unattempted.

Preparing a Test Manual

The revision of the test is not ready for general use until a manual of directions for administering it has been prepared. This manual should be a very explicit step-by-step description of the preparations and procedures involved in giving the test. The importance of uniform procedures has been shown in various studies[8, 13, 22] which reveal differences in test results from station to station, differences apparently due to variations in procedure. The manual should include such topics as suitable conditions for test administration, specific duties of proctors, distributing test booklets, methods of keeping an adequate check on the number of booklets, and seating arrangements, as well as a specific account of the steps to follow in administering the test itself. All words that are intended to be read aloud by the examiner should be printed in boldface type or otherwise set apart typographically in order that they can be found readily at the proper moment in giving the test. The test manual should also contain detailed and specific directions for scoring and for converting raw scores to whatever units are to be used in reporting test results.

14.6 VALIDATING THE TEST

If the test was not validated at the time of its first experimental administration, it is now ready for validation against the criterion which it was designed to predict. The results of the validation study determine which of the various experimental tests that have been tried out will be adopted and put into general use. They also may suggest new hunches with respect to types of tests which may be prepared and tried out.

This experimental administration also furnishes opportunity for a statistical evaluation of the test in its revised form. The reliability of the test may be checked, the items and their distractors may be re-evaluated, and the time limits and test directions may be re-examined. A new item analysis would be particularly important for evaluating items from which extra distractors have been eliminated, since there is always the possibility that the distractors which remain may not function in exactly the same way when a change has been made in their context. After any further revisions have been made, the test is ready to put into general use.

14.7 ESTABLISHING NORMS

14.7.1 Adequacy of the Sample

In establishing norms for a test, the most important consideration, so far as early planning is concerned, is the adequacy of the sample. The sample should be large, in order to avoid sampling fluctuations. It is difficult to state a minimum number, because this would depend to some extent on the number available; but upwards of 2,000 cases would be satisfactory in most instances.

More important than size is the representativeness of the sample. Perhaps the best procedure would be to take *all* cases to which the test would routinely be administered during a particular interval of time (say one month) during which no selective factors are known to be operating.[7, 11]

14.7.2 How to Report Scores

The next problem to be considered is the type of measure which should be used in reporting scores. A desirable measure should possess the following characteristics.

1. It should be easy for the layman to interpret.
2. It should have a standard interpretation, regardless of what particular test is involved.
3. It should not be based on units which are obviously unequal (i.e., a difference of five points near the end of the distribution should mean the same as a difference of five points near the middle of the distribution).
4. It should not be unduly influenced by minor irregularities in the shape of the distribution and should be relatively stable over a period of time.
5. It should involve the use of not more than two digits, for economical use of space on the punch card.
6. It should lend itself to use with available statistical procedures.

Raw scores obviously are unsatisfactory for reporting test scores, since they have no standard interpretation and cannot be directly compared from test to test.

Percentile rank is sometimes used to indicate the relative position of particular scores in the distribution. Percentile rank has the advantage of easy interpretation by the layman, but its disadvantages are too numerous to justify its use: it has unequal units, it is influenced by minor irregularities in the shape of the distribution, and it does not possess arithmetical characteristics which permit its use in statistical procedures.

Some form of standard score most nearly meets the specifications outlined above. Its most important disadvantage is its unfamiliarity to the layman, making interpretation difficult. There is nothing intrinsically difficult, however, in the concept of standard score, and most officers, with a slight amount of indoctrination, can readily learn to use these scores. Standard scores satisfy the requirement of equal units well enough for all practical purposes, provided the score distributions do not depart too markedly from the form of the normal curve, and they do lend themselves to statistical manipulation. By using standard score units based on a mean of 50 and a standard deviation of 10, all standard scores can be expressed in two-digit numbers. The Navy standard scores which are now in use are of exactly this type. Army standard scores, which have a mean of 100 and a standard deviation of 20, require three-digit numbers for reporting half the scores but possess all the other advantages listed. The experience of World War II demonstrated that standard scores can be used by nontechnical officers.

Preparation of Conversion Tables

Conversion tables should be prepared, after the test has been scored, for the purpose of affording a convenient method of transmitting raw scores to standard scores. The conversion table should consist of a listing of all obtainable raw scores, each of which is paired with its corresponding standard score. The conversion table should be included with the manual of directions for administering and scoring the test.

A Supplementary Method of Reporting Scores

For the benefit of officers who are untrained in statistical procedures yet who will need to

[redacted]

make use of test scores to assist in classification problems, it may be desirable to furnish a supplementary method of reporting scores which is simpler and coarser than the standard scores. The method might consist of reducing the scale to about five categories which are defined in terms of percentage of the population at various regions along the continuum. The Enlisted Classification Section of the Bureau of Naval Personnel has included in its recommendations for a postwar classification program the suggestion that such a scale consist of five categories, as follows.

1—top 7 per cent
2—next 24 per cent
3—middle 38 per cent
4—next 24 per cent
5—bottom 7 per cent

Such a scheme used in conjunction with standard scores should come very near to satisfying the criteria listed above. It is important that standard scores be included in each man's record, however, because of their potential value for research studies. A five-point scale would be very inadequate for correlational studies, even if all categories were represented in a particular sample; and it is very likely that, because of selection, only two or three of the categories would be represented in a particular group.

14.8 CONTINUATION OF TEST DEVELOPMENT

The work of test development cannot be considered completed at this point, except in a very restricted sense. A test constructed according to the procedures outlined in this chapter will be a good test at the time of construction; but there is no guarantee that it will continue to be a good test as time passes. Service requirements may change as new equipment and tactics are developed, and the content of training programs may be modified accordingly. As a war progresses, the quality of the available manpower may also change. For these and various other reasons, a continuing program of research is necessary.

Tests should be re-evaluated periodically, particularly from the point of view of validity. Norms should be checked occasionally in order to be aware of any shifts which may occur in the character of the population being tested. The search for improved testing methods and more valid types of test items should be a continuing process. Only through such a continuing program of research can the tests be kept up to their maximum value in classification of military personnel where the importance of tests in maintaining efficiency is recognized.

[redacted]

Chapter 15

CONTRIBUTIONS TO TEST THEORY AND TEST ADMINISTRATION

By *Dael Wolfle*[a]

Summary

An investigation of the accuracy with which aptitude tests were being scored in 1942 in the Navy revealed frequent and large errors. Methods of controlling and checking scoring procedures were devised. Subsequent investigations showed that errors were greatly reduced in frequency and in size.

Several innovations in test construction practice were tried out. They included:

1. A simple method of measuring the premium which a test places on speed.

2. A method of estimating the validity of a test for a school population without waiting for training and performance measures to accumulate after giving the test to the population. Instead, the test is given to a graduating class and their test scores are correlated with final grades.

3. A method for increasing the validity of a test by selecting items in terms of their correlation with an external criterion such as success at the school, instead of correlation with the total-test score.

15.1 INTRODUCTION

At the start of the war the routine procedures used in scoring tests administered in a number of Navy installations were seriously at fault and resulted in frequent and large errors in the test scores reported. Improvements in the mechanics of test scoring are described in Section 15.2.

The need for speed in constructing the tests of the U. S. Navy Basic Classification Test Battery (Chapter 2) and the tests for selection of officers (Chapter 3) left little time for fundamental research or development in test construction or the theory of mental measurement. It was, however, possible to try out, on a fairly large scale, some ideas which had not been thoroughly tested before the war. Sections 15.3 to 15.5 describe the applications of these ideas.

15.2 IMPROVEMENTS IN THE ADMINISTRATION AND SCORING OF NAVY TESTS

Before beginning to develop tests for the Navy Basic Battery or the Officer Qualification Test, detailed analyses were made of the tests in use by the Navy in 1942. This preliminary work included a study of the accuracy with which test papers were being scored at the four naval training stations located at Great Lakes, Illinois; San Diego, California; Newport, Rhode Island; and Norfolk, Virginia.[1]

Very large errors were found in the scoring of some papers. In the GCT, scoring errors ran as high as 33 points, almost one-third of the total score. For the Code Test of 78 items, errors went up to 40 points, or over half the total score. For the MAT with a total score of 324, 1 paper was in error by 131 points and 8 were in error by more than 50 points. On the Arithmetic, English, and Spelling tests, each with 100 points perfect score, scoring errors of as much as 40 points were found. Such large errors, even though infrequent, are serious when individual assignments are made on the basis of test score.

The most accurately scored examination at the most accurate station showed an average error of .06 points on a test containing 29 items. Of the 220 papers checked, 207 were correctly scored.

The most inaccurately scored examination at the most inaccurate station showed an average error of 11.41 points on a test on which the perfect score was 342. Of the 204 papers checked, only 6 were correctly scored.

The best of these records was three times as

[a] This chapter is based upon the work of Project N-106.

bad as the functioning accuracy level of the College Entrance Examination Board or the Educational Records Bureau (two professional testing organizations). The worst record showed over 50 times as large an average error as the maximum ever allowed by these testing organizations.

The fact that Army tests were being scored with much greater accuracy than Navy tests indicated that the fault was not a necessary consequence of using enlisted personnel as test scorers.

In order to reduce the number of scoring errors, instructions for a standardized scoring procedure were written. This procedure is outlined schematically in Figure 1.

The packet slip shown in Figure 2 was devised as a record of the progress of each packet of answer sheets through the scoring routine and as a means of fixing responsibility for such errors as occurred.

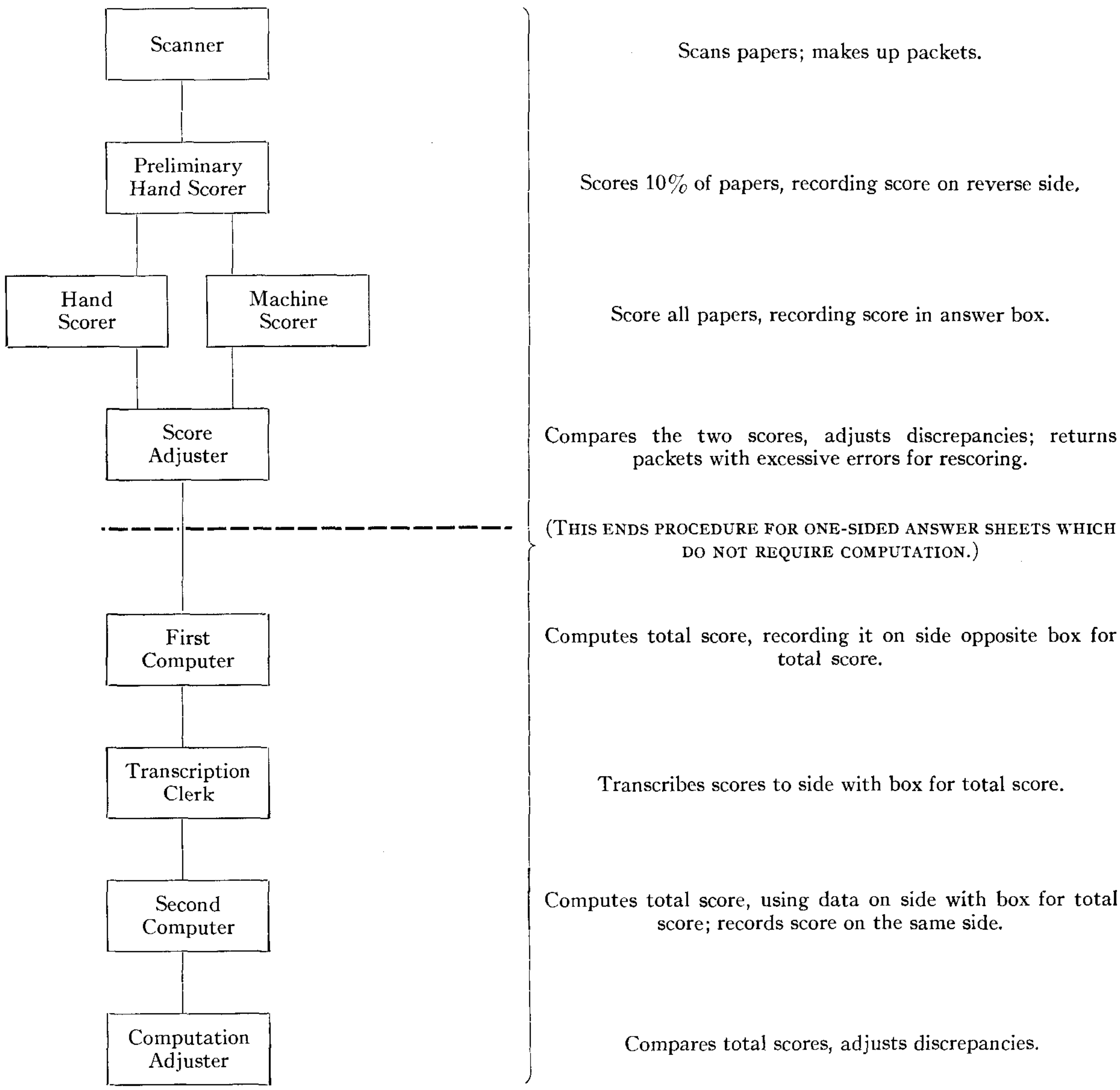


FIGURE 1. Suggested scoring procedure for naval training stations.

Occasional checks made later showed that the adoption of these procedures by the Navy led to a great decrease in the seriousness of scoring

[redacted]

errors. Errors continued to be made with somewhat greater frequency than would be tolerated by a professional testing organization, but they were fewer in number and smaller in size than those found prior to the introduction of the standard scoring procedures.

At the same time, ease of test administration and scheduling requires fixed time limits. If the assigned time limits prove to be too short, what was originally intended to be a power test becomes a speed test. It is desirable, therefore, to have available a method of measuring the

Number of papers	Packet number ....
Packet to be:	Name of test:
Hand-scored	M.A.T. ....
Machine scored....................	G.C.T. ....
	Code ....
	A., E., and S. ....

	Errors	Comments
Scanned and counted by	....	....
Sample of .... papers		
hand-scored by	....	....
Page 1 scored by	....	....
Page 2 scored by	....	....
Page 1 adjusted by	....	....
Page 2 adjusted by	....	....
First computation of total score by	....	....
Transcribed by	....	....
Second computation of total score by	....	....
Total score adjusted by	....	....

APPROVED: ....

Selection Officer

FIGURE 2. Packet slip.

15.3

MEASURING THE PREMIUM PLACED ON SPEED IN A TEST

Tests which grant a high premium for speed or impose a heavy penalty for slowness are commonly called speed tests. In tests with extremely short time limits, speed is likely to play a major role in determining test scores; in tests with long time limits, speed is likely to have little or no effect upon test scores. For many testing purposes, it is generally considered desirable to emphasize power rather than speed. premium which any test, due to its time limits, places on speed of work.

The basic assumption in the development of this measure[4] was that differences in the speed requirements of tests may be measured by differences in the *proportion of individuals who fail to attempt the final item of each test.* The proportion of individuals failing to attempt the final item may be symbolized by S.

$$S = 1.00 - \frac{N_f}{N_i}$$

[redacted]

where N_f = number of persons attempting to answer the *final* item, and N_i = number of persons attempting to answer the *initial* item. N_f/N_i = the proportion of individuals who *attempted* the final item; subtracting from 1.00 gives the proportion who *did not attempt* the final item.

S may be thought of either as a measure of the *premium for speed* or the *penalty for slowness*. S may range in value from .00 to 1.00—the value .00 indicating the absence of premium for speed, and 1.00 indicating a heavy premium for speed. For example, if time is unlimited, so that everyone completes the test, $N_f = N_i$. S then equals zero ($S = 1.00 - 1.00$). If the time limit on another test is short, so that only half the men attempt the final item, N_f is only half as large as N_i, and S equals .50. The higher the value of S, the greater premium there is on speed.

TABLE 1. Premium for speed, as shown by the value of S* (national sample, N=500).

Group of items (test or subtest)	N_i	N_f	S* (Premium for speed)
GENERAL CLASSIFICATION TEST:			
Sentence completion	500	208	.58
Opposites	500	364	.27
Analogies	500	443	.11
READING	500	310	.38
ARITHMETICAL REASONING	500	220	.56
MECHANICAL APTITUDE TEST:			
Block counting	499	106	.79
Mechanical comprehension	500	338	.32
Surface development	489	95	.81
MECHANICAL KNOWLEDGE TEST:			
Elect.-pictorial items (Nos. 1-25)	500	499	.00
Mech.-pictorial items (Nos. 26-55)	500	421	.16
Elect.-verbal items (Nos. 56-90)	500	379	.24
Mech.-verbal items (Nos. 91-135)	500	296	.41

*S is the proportion of individuals who failed to attempt the final item in a test or subtest. Its value is determined from the formula $S = 1.00 - N_f/N_i$ in which N_f is the number who attempted the final item and N_i the number who attempted the initial item on the test. The higher S is, the more the test score was influenced by speed.

S is based on only two items in a test, the first and the last. A more refined, and probably more valid, measure would take into account the number of persons attempting to answer each individual item. Comparisons between the values of S for each test and subtest in the U. S. Navy Basic Classification Test Battery and the number of people attempting each item on each subtest indicate that S is a satisfactory measure.

Table 1 gives the value of S for each test and subtest in the Basic Battery. The use of S clearly shows differences among these tests, all of which were originally intended primarily as power tests.

Deciding whether the value of S is too high, too low, or appropriate for a given test depends, first of all, on whether the test was intended to be a speed test or a power test. For power tests, S should be rather low. There is no very good basis for stating any exact limits. Perhaps the range from .10 to .35 can be suggested as desirable for power tests. Judged by this standard, half the tests, or subtests, listed in Table 1 are found to have time limits which are too short.

If the items toward the end of a test are quite difficult, it makes little difference in test score whether the slower, and usually poorer, individuals attempt them or not. They miss them even when they have time to answer them. The test can therefore be allotted a shorter time, if the final items are very difficult, than would be desirable in a test where all items are of more nearly equal difficulty.

If a test is to be administered by different individuals under different conditions, it is well to err on the side of an excessive time limit (a low value of S) rather than an inadequate one (a high S). The higher the value of S, the more sensitive the test is to irregularities of timing and to unintentional differences among examiners in their emphasis on speed versus accuracy.

The use of S is suggested as an easy method of determining the premium placed on speed by any test and as a basis for determining whether or not appropriate time limits have been established. As experience develops, the meaning of smaller differences in S will become clear. It will then become necessary to refine the measure itself.

[redacted]

15.4 DETERMINING THE VALIDITY OF A TEST

One of the most commonly employed procedures for determining the validity of a test involves the following steps: (1) administering the test to a group of Service school candidates; (2) waiting until the men have completed their Service school training; and then (3) correlating test scores with final achievement grades in the school. Obviously, this procedure demands that one wait as long as three or four months before it becomes possible to make a statement concerning the validity of the test in question.

An alternative procedure,[3] intended to shorten the time necessary to secure validity data, was tried out in developing the Arithmetical Computation Test (AC) described in Chapter 2. It consisted of the following steps:

1. Administering the test to be validated (AC in this case) to classes at the time of graduation and to other classes just entering training.

2. Correlating AC scores with final school grades for the graduating classes. This gave validity coefficients but not necessarily the same ones which would have been obtained had AC been given before training started, for the training itself might have had a differential effect on the distribution of AC scores. In order to determine whether the obtained coefficients were reasonable estimates of those which would have been found by the usual technique, steps 3, 4, and 5 were carried out.

3. Securing from the Q-cards scores on the Basic Battery tests for both entering and graduating classes.

4. Determining regression coefficients, by the usual formula: $b_{12} = r_{12}\left(\frac{\sigma_1}{\sigma_2}\right)$, of AC on GCT and AR, of GCT on AR, and of AR on GCT for both entering and graduating classes.

5. Comparing the regressions of AC on GCT and on AR for graduating classes with those for entering classes and comparing the differences obtained with differences in the regression of GCT on AR and AR on GCT. These last two regressions involved tests which both entering and graduating classes had had prior to their entry into school; differential effects of training therefore could have had no influence on the regressions. Variability in these regressions could be attributed to sampling differences and could be used as a standard for interpreting any differences between entering and graduating class regressions of AC on GCT and on AR.

These comparisons were made for nine different Service schools. None of the differences between regression coefficients of AC on AR or on GCT for entering and for graduating classes approached statistical significance. It was therefore concluded that the validity coefficients obtained for the classes tested at the time of graduation were approximately the same as those which would have been obtained for the same classes had they been tested at entry in the usual manner.

After the classes to which the AC test was given at the time of entry had graduated, it was possible to use their school grades to validate the AC in the usual manner.[5] It was thus possible to compare the estimated validity of the AC with the validity as determined by the usual technique.

The comparison is shown in Table 2. In only two classes did the two validity coefficients differ by more than .10. The mean of the differences was .001. This difference was smaller than the differences in validity coefficients for most of the Basic Battery tests, for which the time of administration was not a variable. (Data for GCT are given in Table 2; comparable data for all Basic Battery tests are given in reference 5.) It was concluded that the differences in AC validity between classes tested at graduation and those tested at entry were due to differences between the samples and not to the influence of training.

This method is recommended for further trial in other situations. If it continues to work satisfactorily, it will be of considerable value in speeding up the validation of a new test.

15.5 SELECTION OF TEST ITEMS BY CORRELATION WITH AN EXTERNAL CRITERION

The criterion usually used in the evaluation of a test item is the total score on the test or

[redacted]

subtest which includes the item. The criterion here is an internal one. Using an *internal criterion* emphasizes test reliability and tends to secure subtests which are homogeneous in content.

It is sometimes desirable to put greater emphasis on test validity than does the usual method of item analysis. In these cases the same *external criterion,* for example school grades, should be used for item analysis that is used for determining the validity of the total test. A method for conducting this type of item analysis has been suggested by Horst[6] and modified by Applied Psychology Panel Project N-106.[2]

TABLE 2. Comparisons of validity coefficients of AC test given at time of entry and at time of graduation to men in six kinds of Service schools.

Service school	*N*		AC			GCT		
	Grad.	Ent.	Grad.	Ent.	Diff.	Grad.	Ent.	Diff.
B. Eng	214	188	.69	.43	−.26	.54	.31	−.23
EM	67	199	.50	.48	−.02	.01	.45	.44
	88	77	.41	.59	.18	.45	.55	.10
FC	125	117	.38	.38	.00	.46	.30	−.16
	73	67	.39	.29	−.10	.35	.25	−.10
QM	48	137	.33	.39	.06	.21	.46	.25
	63	100	.42	.49	.07	.27	.37	.10
SC and Bkr	84	72	.34	.40	.06	.11	.11	.00
SK	115	149	.39	.37	−.02	.38	.44	.06
	111	63	.40	.44	.04	.50	.16	−.34
Mean difference					.001			.012
Mean absolute difference					.081			.178

In attempting to select a set of items such that the correlation with a criterion will be maximized, it is desirable to select individual items which correlate high with the *external criterion* (grades) and relatively low with the *internal criterion* (total-test score). However, such a statement does not take the difficulty of the item into consideration. If item difficulty is included in a theoretical analysis of this problem, it can be shown that it is desirable to have $r_{ic}\sigma_i$ for all the selected items as large as possible and $r_{it}\sigma_i$ as small as possible, where

$\sigma_i = \sqrt{p_i q_i}$, which is the standard deviation of the item;

p_i = the proportion of the population answering an item correctly;

q_i = the proportion of the population answering the item incorrectly;

r_{it} = the point-biserial correlation of the item with total-test score; and

r_{ic} = the point-biserial correlation of the item with the external criterion.

Note: These five statistics are computed from all cases taking the test and not just from the cases attempting each item. That is, a person who records no answer for the last 10 items in a test is counted as having answered those items *incorrectly.*

By this procedure the validity of the test is increased at the expense of decreasing its homgeneity as indicated by the split-half reliability coefficient.

The procedure followed was to compute $r_{it}\sigma_i$, which is a "reliability" index, and $r_{ic}\sigma_i$, which is a "validity" index, for each item; to plot these values, with the validity index on the ordinate and the reliability index on the

abscissa for each of the 104 items in the Mechanical Comprehension Test—OQT O-2. Items lying above and to the left of a diagonal line through the origin were then selected for inclusion in the test. This line was moved down until it included enough items to make up two 30-item tests, called forms 2W and 3W. In order to make the items in the two tests match fairly well with respect to item content, a few items above this line were rejected. The plot is shown in Figure 3. The items selected for

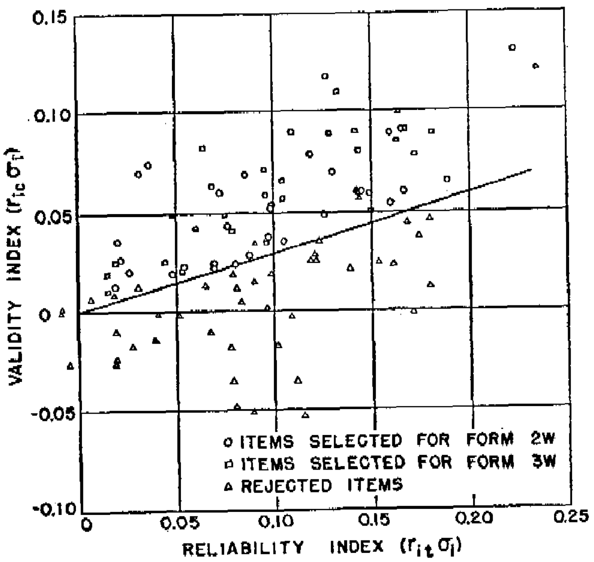


FIGURE 3. Item selection plot for the Mechanical Comprehension Test—OQT O-2.

form 2W of the test are indicated by circles; the items selected for form 3W by squares; and the rejected items by triangles.

The data illustrated in Figure 3 are taken from an analysis of the mechanical comprehension items in the experimental form of the Officer Qualification Test (OQT). Forms 2 and 3 of the OQT were developed from this experimental form (see Chapter 3) but not by means of the method being described, that of correlation with an external criterion. Instead, the items actually included in Forms 2 and 3 of the OQT were selected by the more usual procedure of correlation with an internal criterion. A comparison of the sets of items selected by the two methods showed much overlapping. Of the 60 items actually included in Forms 2 and 3, 40 were also chosen by correlating with an external criterion; 20 items included by correlating with an internal criterion would not have been chosen by correlating with an external criterion.

A sample of the tabulated statistical information computed for each item and used in selecting items on the basis of an external criterion is given in Table 3.

TABLE 3. Item analysis data for items validated against an external criterion.

Item No. in Form O-2	OQT Form	r_{it}	r_{ic}	$r_{it}\sigma_i$	$r_{ic}\sigma_i$	pq	p	N_t
39		.10	.20	.0361	.0745	.2346	.62	418
15	2	.53	.43	.0722	.0607	.0668	.93	418
7	2	.47	.20	.1425	.0606	.0177	.77	418
86	2	.29	−.16	.1149	−.0534	.2238	.34	405

Item 39 was an item of approximately average difficulty ($p = .62$) which did not correlate very well with the other items in the test ($r_{it} = .10$). It was not included in the final forms of the OQT. Its correlation with indoctrination school grades (the external criterion) was somewhat higher than its correlation with the other items, so it would have been included had the external criterion of item selection been used.

Item 15 was an easy item which showed substantial correlations with both criteria. The use of either criterion would dictate its inclusion in the final test.

Items 7 and 86 both showed appreciable correlations with total-test score and both were included in Form 2 of the OQT. One was somewhat below and the other somewhat above average difficulty. When correlated with indoctrination school grades both showed a low, and one a negative, correlation with the criterion. Both would have been rejected by the external criterion method.

Note that items 39 and 7 both had an r_{ic} of .20. But one was somewhat more difficult than the other. Item 39, for which $r_{ic}\sigma_i$ was larger than $r_{it}\sigma_i$, was acceptable on the external criterion basis, while item 7, for which $r_{ic}\sigma_i$ was less than $r_{it}\sigma_i$, was not.

The use of an external criterion for item selection ought to increase the validity of the resulting test over that achieved by using an internal criterion. The available evidence indicates that it does. Forms 2 and 3 of the OQT as actually made up, using the customary internal criterion, had validity coefficients of .35 and .38. Validity coefficients of .46 and .48, for the two forms (2W and 3W) in which items were selected on the basis of an external criterion, were obtained by rescoring the original test papers using only the 60 items found in this test and computing the correlation of the new scores with indoctrination school grades. These coefficients would no doubt be lowered if the tests were used on a new population, but there is no satisfactory method available for estimating the amount of shrinkage. An empirical trial with a new sample was not feasible.

Continued study of the value of this method of selecting items by using it in a number of trial situations is strongly recommended.

GLOSSARY

LABORATORY NAME ABBREVIATIONS

BROWN. Brown University, Providence, Rhode Island.

C.E.E.B. College Entrance Examination Board, Princeton, New Jersey.

HARVARD. Harvard University, Cambridge, Massachusetts.

N.A.S. National Academy of Sciences, Washington 25, D. C.

N.Y.U. New York University, New York City.

PENN. University of Pennsylvania, Philadelphia, Pennsylvania.

PENN. STATE. Pennsylvania State College, State College, Pennsylvania.

PRINCETON. Princeton University, Princeton, New Jersey.

PSYCH. CORP. Psychological Corporation, 522 Fifth Avenue, New York 18, New York.

ROCHESTER. University of Rochester, Rochester, New York.

STANFORD. Stanford University, California.

TUFTS. Tufts College, Medford, Massachusetts.

U.S.C. University of Southern California, Los Angeles, California.

WIS. University of Wisconsin, Madison, Wisconsin.

YERKES LAB. Yerkes Laboratories of Primate Biology, Orange Park, Florida.

OTHER ABBREVIATIONS

AA. Antiaircraft.

AAA. Antiaircraft Artillery.

AAF. Army Air Forces or Army Air Field.

AC. Arithmetical Computation Test.

AGCT. Army General Classification Test.

AGO. Adjutant General's Office.

AI. Airborne Interception.

APP. Applied Psychology Panel.

AR. Arithmetical Reasoning Test.

ARC-1. Army Radio Code Aptitude Test (same as SOR).

AT. Arithmetic Test.

BB. Battleship.

BUPERS. Bureau of Naval Personnel.

CHL. Chain Home Link.

CIC. Combat Information Center.

COMINCH. Commander in Chief, United States Navy.

COTCLANT. Commander, Operational Training Command, United States Atlantic Fleet.

COTCPAC. Commander, Operational Training Command, United States Pacific Fleet.

CRT. Code Receiving Tests.

DD. Destroyer.

DE. Destroyer Escort.

ESF. Equivalent square feet.

EW. Early Warning.

FC. Fire Controlman.

FORM 20. War Department enlisted men's personal record card.

GCI. Ground Controlled Interception.

GCT. General Classification Test (Navy).

JL. A shipboard telephone circuit.

LCVP. Landing Craft, Vehicles and Personnel.

MAT. Mechanical Aptitude Test.

MK-E. Mechanical Knowledge, Electrical.

MK-M. Mechanical Knowledge, Mechanical.

N. Number.

N_t. The number of men in a group who attempted to answer a particular test item.

NAVPERS. Bureau of Naval Personnel publication number.

NDRC. National Defense Research Committee.

NRC. National Research Council.

NROTC. Naval Reserve Officers' Training Corps.

NTS. Naval Training Station.

NTSCH. Naval Training School.

OCS. Officer Candidate School.

OQT. Officer Qualification Test.

OSRD. Office of Scientific Research and Development.

PI. Personal Inventory.

PPI. Plan Position Indicator.

Q-CARD. Navy Enlisted Men's Qualification Card.

R. Reading Test.

r. Correlation coefficient.

r_{bis}. Biserial correlation coefficient.

S. A measure of the premium placed on speed by a test.

σ. Standard deviation.

SCCAT. Signal Corps Code Aptitude Test.

S.D. Standard deviation.

SOR. Speed of Response Test of Code Aptitude.

UOE. Unit of Error.

V-12. A Navy college training program.

WD. War Department.

BIBLIOGRAPHY

CHAPTER 1

1. *The Use of Test Scores and Quality-Classification Ratings in Predicting Success in Electrician's Mates Schools,* Herbert S. Conrad and G. A. Satter, OSRD 5667, CEEB,[a] Sept. 13, 1945. APP-420-M2

2. *Report of Observers, ETO, 11 March to 21 April 1945,* Willard Machle and Frederick S. Brackett, Headquarters ETO, U. S. Army, APO 887, to Office of the Surgeon General, Washington, D. C., Apr. 27, 1945.

CHAPTER 2

1. *Recommendations on Selection, Classification, and Training of Service Personnel,* Memorandum to Chief of the Bureau of Naval Personnel, Washington, D. C., from the Committee on Service Personnel—Selection and Training, National Research Council, July 25, 1942.

2. *Naval Aptitude Tests: A. Reliability; B. Scoring Accuracy; C. Scoring Procedures,* Harold Gulliksen, OSRD 1127, CEEB, Dec. 10, 1942. APP-322-M1

3. *Preparation of the United States Navy General Classification Test—Form 1, and the United States Navy Tests of Reading and Arithmetical Reasoning, Form 1,* H. Gulliksen, Project N-106, Memorandum 7, CEEB, June 23, 1943. APP-327-M1

4. *Averages, Standard Deviations, and Intercorrelations of Navy Aptitude Tests,* Harold Gulliksen, Herbert S. Conrad, and Norman Frederiksen, OSRD 1536, CEEB, June 7, 1943. APP-322-M2

5. *Suggestions for the Revision of the United States Navy Mechanical Aptitude Test, Form T,* H. Gulliksen, Project N-106, Memorandum 2, Jan. 18, 1943. APP-325-M1

6. *Factor Analysis of the New United States Navy Basic Classification Test Battery,* D. A. Peterson, OSRD 3004, CEEB, Sept. 29, 1943. APP-412-M1

7. *Item Analysis of Navy Aptitude Tests,* Herbert S. Conrad, OSRD 3039, CEEB, Dec. 30, 1943. APP-322-M3

8. *Validity of Navy Aptitude Tests in Service Schools at the Great Lakes Naval Training Station,* Norman Frederiksen, OSRD 3245, CEEB, Jan. 31, 1944. APP-322-M4

9. *Statistical Analysis of the Mechanical Knowledge Test,* Herbert S. Conrad, OSRD 3246, CEEB, Jan. 28, 1944. APP-325-M3

10. *Validity of an Experimental Battery of Aptitude Tests at the Ordnance and Gunnery Schools, Washington Navy Yard,* Norman Frederiksen, OSRD 3619, CEEB, Apr. 29, 1944. APP-322-M5

11. *Selection of Items for the US Navy General Classification Test, Form 2, and the US Navy Tests of Reading and Arithmetical Reasoning, Form 2,* G. A. Satter, OSRD 3756, CEEB, June 8, 1944. APP-412-M3

12. *The Preparation of Norms for the Fleet Edition of the General Classification Test,* D. A. Peterson, OSRD 4242B, CEEB, Oct. 10, 1944. APP-414-M1

13. *The Construction and Validation of an Arithmetical Computation Test,* G. A. Satter and Norman Frederiksen, OSRD 4556, CEEB, Jan. 8, 1945. APP-411-M2

14. *A Statistical Evaluation of the Basic Classification Test Battery, Form 1,* Herbert S. Conrad, OSRD 4636, CEEB, May 14, 1945. APP-411-M3

15. *A Further Study of the Validity of the Arithmetical Computation Test,* Norman Frederiksen, OSRD 5302, CEEB, July 3, 1945. APP-411-M4

16. *Summary Report on Research and Development of the Navy's Aptitude Testing Program,* Herbert S. Conrad, OSRD 6110, CEEB, Oct. 31, 1945. APP-322-M6

17. *Statistical Method,* Truman L. Kelley, Macmillan Co., New York, N. Y., 1923, Formula 186, p. 225.

18. *Enlisted Classification Bulletin,* Navy Bureau of Personnel, September 1945, p. 8.

19. *Predictive Efficiency of the Navy Basic Test Battery at Naval Training School (Gunners Mates —Electrical Hydraulics), Navy Yard, Washington, D. C.,* Project 224, Navy Bureau of Personnel, Test and Research Section, July 1, 1945.

20. "Psychological Test Construction and Research in the Bureau of Naval Personnel: Development of the Basic Test Battery for Enlisted Personnel," Staff of the Test and Research Section, *Psychological Bulletin,* Vol. 42, 1945, pp. 561-571.

21. "Psychological Test Construction and Research in the Bureau of Naval Personnel: Validity of the Basic Test Battery, Form 1, for Selection for Ten Types of Elementary Naval Training Schools," Staff of the Test and Research Section, *Psychological Bulletin,* Vol. 42, 1945, pp. 638-644.

---
[a] College Entrance Examination Board.

[redacted]

CHAPTER 3

1. *Preparation of a Preliminary Form of an Officers' Selection Test*, H. Gulliksen, Project N-106, Memorandum 1, CEEB, Dec. 11, 1942. APP-319-M1

2. *Analysis of the N.R.O.T.C. Selective Examination (Form C) and Suggestions for Its Revision*, Harold Gulliksen, OSRD 1290, CEEB, Feb. 15, 1943. APP-411-M1

3. *Preparation of the United States Navy Officer Qualification Test, Form 1*, Harold Gulliksen, OSRD 1273, CEEB, Mar. 12, 1943. APP-319-M2

4. *Preparation of Norms for the Officer Qualification Test, Form 1*, Norman Frederiksen, Project N-106, Memorandum 4, CEEB, May 28, 1943. APP-319-M3

5. *Comparison of the Officer Qualification Test, Form 1, and the United States Navy Aptitude Test, Form E-2*, Norman Frederiksen, Project N-106, Memorandum 5, CEEB, June 1, 1943.

6. *Preparation of Norms for Women for the Officer Qualification Test, Form 1*, H. Gulliksen, Project N-106, Memorandum 9, CEEB, July 24, 1943. APP-319-M4

7. *Development and Validity of the United States Navy Officer Qualification Test*, Norman Frederiksen and D. A. Peterson, OSRD 3186, CEEB, Jan. 7, 1944. APP-319-M5

8. *Study of Combat Leadership*, W. S. Hunter, Memorandum to General Liaison Officers of the Applied Psychology Panel, NDRC, Mar. 3, 1944.

9. *Report to Applied Psychology Panel on Combat Leadership*, Henry E. Garrett and Ernest M. Ligon, NAS, June 8, 1944.

10. *A Statistical Evaluation of the United States Navy Officers Qualification Test, Forms 2 and 3*, D. A. Peterson, OSRD 3978, CEEB, Aug. 4, 1944. APP-319-M6

11. *Preliminary Summary of Results of the Study of the Effectiveness of the Classification Procedure for Officers, Amphibious Training Command, US Atlantic Fleet*, Kinsley R. Smith and Harold A. Voss, Project N-117, Informal Memorandum 6, Pennsylvania State College, Apr. 12, 1945. APP-512-M5

CHAPTER 4

1. *Standardization and Validation of the Personal Inventory: Psychiatric Criterion*, W. C. Shipley, Florence E. Gray, and Nancy Newbert, OSRD 1606, Brown University, June 24, 1943. APP-321-M1

2. *A Report of Research on Selection Tests at the US Submarine Base, New London, Connecticut*, C. H. Graham, L. A. Riggs, N. R. Bartlett, R. N. Berry, W. S. Verplanck, R. L. Solomon, and C. G. Mueller, OSRD 1770, Brown University, June 28, 1943.

3. *Results Obtained from Testing Recruits with the New London-NDRC Questionnaire at the Newport Naval Training Station*, F. A. Mote, R. N. Berry, and C. H. Graham, OSRD 3040, Brown University, Dec. 6, 1943. APP-324-M1

4. *The Relation of Selection Test Scores to Tank Escape Performance at the Submarine School*, C. H. Graham, F. A. Mote, and R. N. Berry, OEMsr-834, OSRD 3262, Project N-113, Brown University, Jan. 31, 1944. APP-413-M2

5. *Item Analysis and Evaluation of the Scoring Stencil of the Personal Inventory*, W. C. Shipley, Florence E. Gray, and Nancy Newbert, OSRD 3315, Brown University, Feb. 14, 1944. APP-412-M2

6. *Test Scores of Amphibious Forces Personnel at Solomons, Maryland, Contrasted with Those Obtained by Personnel of Other Navy Activities*, C. H. Graham, Project N-113, Memorandum 1, Brown University, Mar. 14, 1944.

7. *The Personal Inventory, Short Form, Format C: Derivation and Preliminary Psychiatric Validation*, W. C. Shipley, Florence E. Gray, and Nancy Newbert, OSRD 3390, Brown University, Mar. 15, 1944. APP-321-M3

8. *The Comparability of Formats A and B of the Personal Inventory*, R. N. Berry, H. J. Leavitt, and F. A. Mote, OSRD 3582, Brown University, Apr. 21, 1944. APP-321-M4

9. *The Personal Inventory, Short Form, Format C: Psychiatric Validation on a Pre-Testing Basis*, W. C. Shipley, Florence E. Gray, and Nancy Newbert, OSRD 3604, Brown University, May 1, 1944. APP-321-M5

10. *A Comparison of Personal Inventory Scores with Service Records One Year After Testing*, Walter C. Shipley, Florence E. Gray, and Nancy Newbert, OSRD 3755, Brown University, June 10, 1944. APP-321-M6

11. *Results From the Long and Short Forms of the Personal Inventory and the General Classification Test*, R. N. Berry, C. H. Graham, and F. A. Mote, OSRD 3962, Brown University, July 31, 1944. APP-321-M7

12. *Final Report in Summary of Research on the Personal Inventory and Other Tests*, W. C. Shipley and C. H. Graham, OSRD 3963, Brown University, Aug. 1, 1944. APP-321-M8

13. *Item Differentiation and Derivation of a 50-Item Scoring Key for the Officer Personal Inventory, Form 1*, Kinsley R. Smith and Harold A. Voss, Project N-117, Memorandum 4, Pennsylvania State College, Feb. 8, 1945. APP-412-M5

14. *The Selection of Emotionally Unstable Personnel by Battle Noise Equipment*, Eugene L. Hartley and Dorothea B. Jones, OSRD 4684, Brown University, Feb. 10, 1945. APP-160-M1

15. *An Evaluation of the Personal Inventory for Predicting Success in Parachute School*, G. A. Satter, OSRD 4870, CEEB, Mar. 28, 1945. APP-321-M9

16. *NDRC Battle Noise Equipment, Utilization at Infiltration Range*, Eugene L. Hartley, Project N-113, Informal Memorandum 6, Brown University, Apr. 3, 1945. APP-630-M2

17. *Final Summary of Research on the Use of Battle Noise Equipment*, Eugene L. Hartley and Dorothea B. Jones, OSRD 4931, Brown University, Apr. 12, 1945. APP-630-M3

18. *An Evaluation of the Personal Inventory and Certain Other Measures in the Prediction of Submarine Officers' Evaluations of Enlisted Men*, G. A. Satter, OSRD 5557, CEEB, Sept. 7, 1945. APP-321-M10

19. *Intercorrelations of Items of the NDRC Personal Inventory, and Correlations of These Items with Otis Intelligence Scores*, N. R. Bartlett, Bureau of Medicine and Surgery Research Project X-460 (Sub. 112), Medical Research Department, U. S. Submarine Base, New London, Conn., June 2, 1945.

20. *Application of the NDRC Personal Inventory to Marine Corps Officer Candidates: Preliminary Validation*, S. B. Williams and H. J. Leavitt, Bureau of Medicine and Surgery Research Project X-386 (Sub. 75), Camp LeJeune, N. C., Dec. 26, 1944.

21. *Application of the NDRC Personal Inventory to Marine Corps Officer Candidates, Progress Report No. 2: Further Validation*, S. B. Williams, H. J. Leavitt, and G. V. Mendola, Bureau of Medicine and Surgery Research Project X-386 (Sub. 75), Medical Field Research Laboratory, Camp LeJeune, N. C., Nov. 27, 1945.

22. *Application of the NDRC Personal Inventory to Marine Corps Officer Candidates, Progress Report No. 3: Data on Final 671 Candidates*, H. J. Leavitt, S. B. Williams, and N. J. Lipkin, Bureau of Medicine and Surgery Research Project X-386 (Sub. 75), Medical Field Research Laboratory, Camp LeJeune, N. C., Dec. 20, 1945.

23. *The Development of the MFRL Confidential Questionnaire*, S. B. Williams, H. J. Leavitt, and C. L. Blair, Bureau of Medicine and Surgery Research Project X-615 (Sub. 133), Medical Field Research Laboratory, Camp LeJeune, N. C., Dec. 21, 1945.

24. History of AAF Aviation Psychology Program. (Specific information not available.)

CHAPTER 5

1. *Report on an Activity-Preference Test for the Classification of Service Personnel*, Truman L. Kelley, OSRD 4484, Final Report of Project SOS-7, Harvard University, Dec. 21, 1944. APP-521-M2

CHAPTER 6

1. *The Experimental Edition of Code Receiving Tests*, Albert K. Kurtz, Harold G. Seashore, and John M. Willits, OSRD 1314, Psychological Corp., Mar. 29, 1943. APP-317-M1

2. *Instructor's Manual for the Code Receiving Tests*, NDRC Project N-107, Psychological Corp., November 1943. APP-317-M2

3. *Tentative Manual and Guide for Using the Radio Code Test: Speed of Response, Form 1*, NDRC Project N-107, Psychological Corp., Dec. 15, 1943. APP-317-M3

4. *The Prediction of Code Learning Ability*, Albert K. Kurtz, OSRD 4059, Psychological Corp., Aug. 26, 1944. APP-212.1-M1

5. *Experiments in Training Radar Operators in Visual Code Reception*, Irving H. Anderson, Alfred L. Baldwin, Robert S. Daniel, William H. Lichte, Thomas L. McCulloch, Fred McKinney, and Donald B. Lindsley, OSRD 4811, Yerkes Laboratories, Mar. 20, 1945. APP-212.5-M1

6. "Eine Eignungsprufüng für Funkentelegraphisten," R. A. Biegel, *Psychotechnische Zeitschrift*, Vol. 6, April 1931, pp. 41-45.

7. *Report on a Code Aptitude Test*, L. L. Thurstone, University of Chicago, May 1943.

8. *Report on Radio Code Aptitude Tests, PR 4000*, Adjutant General's Office, Technical Section, Classification and Replacement Branch, 1944.

9. *Memorandum on Comparison of Code Aptitude Test (Thurstone) and Radio Code Aptitude Test (Speed of Response)*, Pers-4122b-do, U. S. Navy, June 7, 1944.

CHAPTER 7

1. *Instruction Manual for Oscilloscope Operator Tests: Experimental Edition*, Donald B. Lindsley, OSRD 1811, Yerkes Laboratories, Sept. 1, 1943. APP-318-M2

2. *Oscilloscope Operator Tests: Experimental Edition*, Donald B. Lindsley, OSRD 1812, Yerkes Laboratories, Sept. 1, 1943. APP-318-M1

[redacted]

3. *Preliminary Report of Results from Oscilloscope Operator Tests: Experimental Edition,* Donald B. Lindsley, OSRD 1813, Yerkes Laboratories, Sept. 1, 1943. APP-318-M3

4. *Proficiency Measures for Classification and Assignment of Operators of ASV Radar Equipment,* Irving H. Anderson and others, OSRD 2012, Yerkes Laboratories, Sept. 29, 1943. APP-318.1-M1

5. *Development of Proficiency Measures for Classification and Assignment of Operators of ASV Radar Equipment,* Irving H. Anderson and others, OSRD 2013, Yerkes Laboratories, Sept. 29, 1943. APP-318.1-M2

6. *Vision as Related to Proficiency in Oscilloscope Operation,* Irving H. Anderson and others, OSRD 3409, Yerkes Laboratories, Feb. 24, 1944. APP-318-M5

7. *A Validational Study of Oscilloscope Operator Tests,* Irving H. Anderson and others, OSRD 3712, Yerkes Laboratories, Apr. 24, 1944. APP-318-M6

8. *Report on Selection of Intercept Officer Candidates,* George K. Bennett, NDRC Project N-109, Informal Memorandum 3, Psychological Corp., Jan. 19, 1945. APP-170-M2

9. *Prediction of Success in Training and Measurement of Achievement at Naval Training School (Tactical Radar),* Bureau of Naval Personnel Project 412 Research Report, May 1, 1945.

CHAPTER 8

1. *Selection of Stereoscopic Range Finder Operators,* OSRD 410, Section D-2, Fire Control, NDRC, Progress Report to the Services 22, Feb. 23, 1942. Div. 7-220.14-M16

2. *Description of Instruments for Use in the Selection of Stereoscopic Range Finder Operators,* OSRD 592, Section D-2, Fire Control, NDRC, Report to the Services 30, May 18, 1942. Div. 7-220.14-M4

3. *Manual for Use in the Selection of Stereoscopic Range Finder Operators,* OSRD 721, Section D-2, Fire Control, NDRC, Report to the Services 31, July 7, 1942. Div. 7-220.14-M11

4. *Analysis of Factors Effective in the Use of the Stereoscopic Range and Height Finder,* OSRD 862, Section D-2, Fire Control, NDRC, Report to the Services 32, Sept. 9, 1942. Div. 7-220.213-M3

5. *The Chromatic Dispersion of the Human Eye and Its Possible Influence on Stereoscopic Range Finding,* OSRD 923, Section D-2, Fire Control, NDRC, Report to the Services 35, Oct. 1, 1942. Div. 7-230.1-M2

6. *Supplement to: Manual for Use in the Selection of Stereoscopic Range Finder Operators,* OSRD 1210, Div. 7, Fire Control, NDRC, Report to the Services 49, Feb. 17, 1943. Div. 7-220.14-M15

7. *Sex Differences in Stereoscopic Ranging and in Visual Tracking,* OSRD 1250, Div. 7, Fire Control, NDRC, Report to the Services 52, Mar. 9, 1943. Div. 7-220.34-M2

8. *Manual for the Adjustment and Operation of the Projection Eikonometer,* H. A. Imus, OSRD 1340, Brown University, Mar. 29, 1943. APP-122-M1

9. *Distribution of Measures of Interpupillary Distance,* H. A. Imus, OSRD 1341, Brown University, Mar. 24, 1943. APP-121-M1

10. *The Adjustment of the M2 Trainer for Standard Testing,* H. M. Fisher and H. A. Imus, OSRD 1638, Brown University, June 17, 1943. APP-652-M3

11. *The Relationship between Test Scores Obtained on the Single and Multiple Projection Eikonometers,* H. A. Imus, OSRD 1789, Brown University, Aug. 5, 1943. APP-122-M2

12. *The Selection of Stereoscopic Height Finder Observers,* W. C. Biel, W. J. Brogden, W. F. Dearborn, H. M. Fowler, C. H. Graham, H. A. Imus, W. E. Kappauf, and F. A. Mote, OSRD 1790, Brown University, Aug. 14, 1943. APP-120-M1

13. *The Relationship of Visual Acuity to Acuity of Stereoscopic Vision,* J. K. Adams, H. M. Fowler, and H. A. Imus, OSRD 2087, Brown University, Sept. 15, 1943. APP-121-M2

14. *The Reliability and Precision of the NDRC and Bausch and Lomb Interpupillometers,* J. K. Adams, D. C. Beier, and H. A. Imus, OSRD 3475, University of Wisconsin, Mar. 29, 1944. APP-112.2-M1

15. *Final Report in Summary of Work on the Selection and Training of Stereoscopic Height Finder Observers,* W. C. Biel, W. J. Brogden, C. H. Graham, H. A. Imus, and W. E. Kappauf, OSRD 3773, Brown University, June 14, 1944. APP-120-M2

16. *The Reliability of Two Forms of an Activity Preference Blank,* J. K. Adams and H. M. Fowler, OSRD 3778, University of Wisconsin, June 19, 1944. APP-112.1-M1

17. *The Construction and Calibration of the NDRC Interpupillometer,* D. G. Ellson and H. A. Imus, OSRD 3797, University of Wisconsin, June 19, 1944. APP-112.2-M2

18. *A Test-Retest Reliability Study of the Bausch and Lomb Ortho-Rater with Naval Personnel,* J. K. Adams, D. C. Beier, and H. A. Imus, OSRD 3969, University of Wisconsin, Aug. 1, 1944. APP-111.1-M1

[redacted]

19. *Manual for Use in the Selection of Fire Controlmen (O), (Stereoscopic Rangefinder Operators)*, H. A. Imus, OSRD 4050, University of Wisconsin, Aug. 22, 1944. APP-111-M1

20. *Manual for the Installation and Adjustment of the Multiple Projection Eikonometer*, H. A. Imus, OSRD 4233, University of Wisconsin, Oct. 10, 1944. APP-122-M3

21. *A Follow-up Study of the Efficiency of the Projection Eikonometer Test in Predicting the Performance of Stereoscopic Height Finder Observers*, D. C. Beier, Florence Gray, H. A. Imus, and E. B. Knauft, OSRD 4352, University of Wisconsin, Nov. 21, 1944. APP-122-M4

22. *The Selection of Fire Controlmen (O)—Rangefinder and Radar Operators*, D. C. Beier and Florence Gray, OSRD 4861, University of Wisconsin, Mar. 26, 1945. APP-111-M2

23. *Final Report in Summary of Research on the Selection and Training of Rangefinder and Radar Operators*, W. J. Brogden, OSRD 5358, University of Wisconsin, July 23, 1945. APP-110-M1

24. *Intercorrelations between Scores on the Tufts Stereo-Trainer; the Navy Stereo-Trainer; the Bott Test of Stereo Acuity; the Vectographic Pursuit Apparatus; the Wulfeck Group Test of Stereo Acuity; and Two Tests of General Intelligence*, Project 10, Progress Report to NDRC, Tufts College, June 18, 1942. Div. 7-220.15-M3

25. *The Correlation between the Keystone and Bott Tests of Stereopsis and Influence of Size of Test Stimuli in Judgments of Stereopsis*, Project 10, Progress Report to NDRC, Tufts College, June 18, 1942. Div. 7-220.15-M6

26. *The Test-Retest Reliability Coefficients of the Tufts Stereo-Trainer, the Navy Stereo-Trainer, Mark* II, *and the Bott Test of Stereo Acuity*, Project 10, Progress Report to NDRC, Tufts College, June 18, 1942. Div. 7-220.15-M4

27. *Validation and Standardization of Tests Used in the Selection of Stereoscopic Rangefinder Operators*, Fire Control Research, Progress Report 15 to NDRC, Princeton University, July 13, 1942.

28. *Interrelationships among Seven Tests of Stereoscopic Acuity and the Relationship between Two Tests of Visual Acuity and Two Tests of Phorias*, H. M. Fowler, H. A. Imus, and F. A. Mote, Memorandum 12, Brown University, Mar. 24, 1944. APP-121-M3

29. *GCT and MAT Selection Standards for Fire Controlmen (R)*, W. J. Brogden, Project N-114, Informal Memorandum 1, Apr. 28, 1944. APP-112-M1

CHAPTER 9

1. *How to Use the NDRC Model* III *Dark Adaptometer*, Rochester, N. Y., December 1942. APP-151-M1

2. *A Study of the Prediction of Night Lookout Performance*, Carl H. Wedell, OSRD 3357, Princeton University, Mar. 15, 1944. APP-150-M1

3. *General Recommendations Concerning Lookouts*, Carl H. Wedell, Memorandum to the Applied Psychology Panel, Project N-115, Memorandum 1, Princeton University, Sept. 15, 1944. APP-150-M2

4. *A Study of the Performance of Night Lookouts Aboard Ship*, W. C. H. Prentice, OSRD 4087, Princeton University, Oct. 7, 1944. APP-150-M3

5. *Final Report in Summary of Work on the Selection and Training of Night Lookouts*, C. H. Wedell, OSRD 4342, Princeton University, Nov. 15, 1944. APP-150-M5

CHAPTER 10

1. *A Speech Interview for the Selection of Telephone Talkers*, George K. Bennett and John W. Black, OSRD 1769, Psychological Corp., August 1943. APP-130-M1

2. *A Preliminary Study of the Abilities of Rated Men to Judge Speaking Performance*, John C. Snidecor, OSRD 1823, Psychological Corp., August 1943. APP-218.31-M1

3. *Report on Seminar on Telephone Talking and Composition of a Manual for the Use of Sound-Powered Phones*, John C. Snidecor, NDRC Project N-109, Informal Memorandum 1, Psychological Corp., Sept. 28, 1943. APP-218.3-M1

4. *A Manual for the Training of Interviewers*, John C. Snidecor and Grant Fairbanks, OSRD 1827, Psychological Corp. APP-218.31-M3

5. *A Study in Training Classification Petty Officers to Select Telephone Talkers*, G. K. Bennett and J. W. Black, OSRD 1931, Psychological Corp., November 1943. APP-218.31-M2

6. *Methods of Training Telephone Talkers for Increased Intelligibility*, John C. Snidecor, Louis A. Mallory, and Edward L. Hearsey, OSRD 3178, Psychological Corp., Jan. 28, 1944. APP-218.2-M1

7. *Information Regarding an Analysis of Recordings Made from the Telephone Circuits on Board a Battleship and a Destroyer*, G. K. Bennett and J. W. Black, OSRD 3312, Psychological Corp., February 1944. APP-218.3-M2

8. *Final Report in Summary of Work on the Selection and Training of Telephone Talkers*, Louis A. Mallory and William J. Temple, OSRD 5497, Psychological Corp., Aug. 27, 1945. APP-218.1-M2

[redacted]

CHAPTER 11

1. *Manual of Instruction for the Use of Findex Equipment in Personnel Placement: USS New Jersey,* John M. Stalnaker, Committee on Service Personnel—Selection and Training, National Research Council, June 28, 1943. APP-521-M1

2. *An Analysis of Men Entering the DD Pool, Precommissioning Training Center, Treasure Island, May 29 to July 29, 1944,* Max M. Levin, OSRD 4151, Stanford University, Sept. 20, 1944. APP-513-M1

3. *A Procedure for Sorting the Enlisted Personnel Qualifications Cards,* G. A. Satter and N. C. Kephart, OSRD 4689, CEEB, Jan. 22, 1945. APP-532-M1

4. *The Selectometer, a Classification Guide,* Evan R. Keisler, OSRD 4746, Stanford University, Feb. 24, 1945. APP-521-M3

5. *Comments on the Manual Entitled "Handbook of Interviewing Guides for Destroyer (2200-Ton) Billets, Repair Party, Engine Room, Fire Room"* (with accompanying handbook), Morris S. Viteles and John H. Gorsuch, NDRC Project NR-106, Memorandum 11, University of Pennsylvania, May 14, 1945. APP-511-M2

6. *Classification Data Available to Ships' Officers,* D. A. Peterson, OSRD 5145, CEEB, May 30, 1945. APP-180-M1

7. *A Point Score Method for Evaluating Naval Personnel,* Max M. Levin, Ronald K. Campbell, Evan R. Keisler, and J. Bradley Sonderman, OSRD 5197, Stanford University, June 12, 1945. APP-522-M1

8. *An Experimental Personnel Record System for Shipboard Use,* OSRD 5303, CEEB, July 3, 1945. APP-531-M2

9. *Memorandum on the Construction and Validation of a Work Readiness Test for Distilling Plant Operators,* B. J. Covner, H. A. Voss, and S. M. Wesley, OSRD 5507, University of Pennsylvania, Aug. 29, 1945. APP-316-M1

10. *Memorandum on Constructing Work Readiness Tests,* H. A. Voss and S. M. Wesley, OSRD 5521, University of Pennsylvania, Aug. 31, 1945. APP-328-M1

11. *Final Report in Summary of Research and Development of Classification Aids by NDRC Project N-116a,* Ronald K. Campbell, OSRD 5647, Stanford University, Sept. 11, 1945. APP-520-M1

12. *Handbook of Interviewing Guides for Billets on 2,200-Ton Destroyers: Ship Control, Combat Information, Communication, Damage Control, Engineering, Gunnery* (with accompanying handbook), A. G. Bayroff and R. M. Jones, OSRD 6115, University of Pennsylvania, Oct. 16, 1945. APP-511-M4

13. *Manual of Classification Procedures for Amphibious Training Bases,* Amphibious Training Command, U. S. Atlantic Fleet, Naval Operating Base, Norfolk, Va., Sept. 1, 1944.

CHAPTER 12

1. *A Point Score Method for Evaluating Naval Personnel,* Max M. Levin, Ronald K. Campbell, Evan R. Keisler, and J. Bradley Sonderman, OSRD 5197, Stanford University, June 12, 1945. APP-522-M1

2. *Predicting Success in Service School from the Order of Assignment,* G. A. Satter and Herbert S. Conrad, OSRD 5556, CEEB, Sept. 7, 1945. APP-420-M1

3. *Use of Test Scores and Quality-Classification Ratings in Predicting Success in Electrician's Mates School,* Herbert S. Conrad and G. A. Satter, OSRD 5667, CEEB, Sept. 13, 1945. APP-420-M2

4. *Summary Report on Research and Development of the Navy's Aptitude Testing Program,* Herbert S. Conrad, OSRD 6110, CEEB, Oct. 31, 1945. APP-322-M6

5. *Manual for Processing the Enlisted Personnel Qualifications Card,* U. S. Navy, NavPers 16704, January 1945.

6. *Manual of Procedures for U. S. Naval Training Station Selection Department,* U. S. Navy, NavPers 16710.

CHAPTER 13

1. *A Speech Interview for the Selection of Telephone Talkers,* George K. Bennett and John W. Black, OSRD 1769, Psychological Corp., August 1943. APP-130-M1

2. *Proficiency Measures for Classification and Assignment of Operators of ASV Radar Equipment,* Irving H. Anderson and others, OSRD 2012, Yerkes Laboratories, Sept. 29, 1943. APP-318.1-M1

3. *Memorandum on Service School Grades,* Harold Gulliksen, OSRD 3177, CEEB, Jan. 6, 1944. APP-413-M1

4. *Manual for Use in the Selection of Fire Controlmen (O), (Stereoscopic Rangefinder Operators),* H. A. Imus, OSRD 4050, University of Wisconsin, Aug. 22, 1944. APP-111-M1

5. *The Code Receiving Tests,* Albert K. Kurtz and Harold Seashore, OSRD 3157, Psychological Corp., Sept. 12, 1944. APP-317-M4

6. *A Procedure for Sorting the Enlisted Personnel Qualifications Cards,* G. A. Satter and N. C. Kephart, OSRD 4689, CEEB, Jan. 22, 1945. APP-532-M1

7. *Final Report in Summary of Work on the Selection and Training of Cargo Handling Teams for Combat Laden Vessels,* Floyd L. Ruch, OSRD 5140, USC, May 29, 1945. APP-222-M1

8. *Classification Data Available to Ships' Officers,* D. A. Peterson, OSRD 5145, CEEB, May 30, 1945. APP-180-M1

9. *The Development of Achievement Tests for Gunner's Mates Schools,* Norman Frederiksen and A. E. Monroe, OSRD 5259, CEEB, June 25, 1945. APP-311-M1

10. *An Experimental Personnel Record System for Shipboard Use,* OSRD 5303, CEEB, July 3, 1945. APP-531-M2

11. *Final Report in Summary of Work on the Job Analysis, Qualification, and Placement of Personnel in the Amphibious Force,* Kinsley R. Smith, OSRD 5422, Pennsylvania State College, Aug. 8, 1945. APP-512-M8

12. *Achievement Examinations for Signalman School,* M. Duane Bown, Thelburn L. Engle, and Harold P. Bechtoldt, OSRD 5460, CEEB, Aug. 20, 1945. APP-312-M1

13. *Memorandum on the Construction and Validation of a Work Readiness Test for Distilling Plant Operators,* B. J. Covner, H. A. Voss, and S. M. Wesley, OSRD 5507, University of Pennsylvania, Aug. 29, 1945. APP-316-M1

14. *Development of Achievement Tests for Class "A" Naval Training Schools (Torpedomen),* Charles M. Harsh and Joseph B. Cooper, OSRD 5520, CEEB, Aug. 31, 1945. APP-311-M2

15. *Memorandum on Constructing Work Readiness Tests,* H. A. Voss and S. M. Wesley, OSRD 5521, University of Pennsylvania, Aug. 31, 1945. APP-328-M1

16. *An Evaluation of the Personal Inventory and Certain Other Measures in the Prediction of Submarine Officers' Evaluations of Enlisted Men,* G. A. Satter, OSRD 5557, CEEB, Sept. 7, 1945. APP-321-M10

17. *The Development of Performance Tests for Use in Class "A" Electrical Schools,* Merrell E. Thompson, and Norman Frederiksen, OSRD 5666, CEEB, Sept. 13, 1945. APP-315-M1

18. *Development of Achievement Tests for the Landing Craft School, Coronado, California,* Norman Frederiksen, OSRD 5670, CEEB, Sept. 14, 1945. APP-314-M1

19. *Development of Radio Code Receiving Examinations,* M. Duane Bown and Harold O. Gulliksen, NDRC Project N-106, Memorandum 18, CEEB, Sept. 27, 1945. APP-317-M7

20. *Summary Report on Research and Development of the Navy's Aptitude Testing Program,* Herbert S. Conrad, OSRD 6110, CEEB, Oct. 31, 1945. APP-322-M6

21. *Manual of Procedures for U. S. Naval Classification Centers,* NavPers 15082, Classification and Selection Section, Bureau of Naval Personnel, February 1945.

22. *Are Destroyer Radiomen Ready for Battle?* [U. S. Navy publication. (?)]

23. *Memorandum to Director of Planning and Control,* J. B. Hogle, Pers 6370-HMG-1, Bureau of Naval Personnel, Jan. 3, 1946.

24. *Manual of Navy Enlisted Job Classification,* NavPers 15105, U. S. Navy.

CHAPTER 14

1. *Naval Aptitude Tests: A. Reliability; B. Scoring Accuracy; C. Scoring Procedures,* OSRD 1127, CEEB, Dec. 10, 1942. APP-322-M1

2. *Preparation of a Preliminary Form of an Officers' Selection Test,* NDRC Project N-106, Memorandum 1, CEEB, Dec. 11, 1942. APP-319-M1

3. *Suggestions for the Revision of the United States Navy Mechanical Aptitude Test, Form T,* NDRC Project N-106, Memorandum 2, CEEB, Jan. 18, 1943. APP-325-M1

4. *Analysis of the NROTC Selective Examination, Form C, and Suggestions for Its Revision,* OSRD 1290, CEEB, Feb. 15, 1943. APP-411-M1

5. *Preparation of the United States Navy Officer Qualification Test, Form 1,* OSRD 1273, CEEB, Mar. 12, 1943. APP-319-M2

6. *Preparation of an Experimental Form of a Spatial Relations Test,* Norman Frederiksen, NDRC Project N-106, Memorandum 3, CEEB, May 27, 1943. APP-323-M1

7. *Preparation of Norms for the Officer Qualification Test, Form 1,* Norman Frederiksen, NDRC Project N-106, Memorandum 4, CEEB, May 28, 1943. APP-319-M3

8. *Averages, Standard Deviations, and Intercorrelations of Navy Aptitude Tests,* Harold Gulliksen, Herbert S. Conrad, and Norman Frederiksen, OSRD 1536, CEEB, June 7, 1943. APP-322-M2

9. *Preparation of the United States Navy General Classification Test, Form 1, and the United States Navy Tests of Reading and Arithmetical Reasoning, Form 1,* Norman Frederiksen, NDRC Project N-106, Memorandum 7, CEEB, June 23, 1943. APP-327-M1

10. *Preparation of Experimental Forms of the United States Navy General Classification Test and the United States Navy Tests of Reading and Arithmetical Reasoning,* Norman Frederiksen, NDRC Project N-106, Memorandum 8, CEEB, June 28, 1943.

11. *Preparation of Norms for Women for the Officer Qualification Test, Form 1,* NDRC Project N-106, Memorandum 9, CEEB, July 24, 1943. APP-319-M4

12. *Factor Analysis of the New United States Navy Basic Classification Test Battery,* D. A. Peterson, OSRD 3004, CEEB, Sept. 29, 1943. APP-412-M1

13. *Item Analysis of Navy Aptitude Tests,* Herbert S. Conrad, OSRD 3039, CEEB, Dec. 30, 1943. APP-322-M3

14. *Development and Validity of the United States Navy Officer Qualification Test,* Norman Frederiksen and D. A. Peterson, OSRD 3186, CEEB, Jan. 7, 1944. APP-319-M5

15. *Selection of Test Items by Correlation with an External Criterion, as Applied to the Mechanical Comprehension Test—OQT-O-2,* Harold Gulliksen, OSRD 3187, CEEB, Jan. 8, 1944. APP-325-M2

16. *Statistical Analysis of the Mechanical Knowledge Test,* Herbert S. Conrad, OSRD 3246, CEEB, Jan. 28, 1944. APP-325-M3

17. *Selection of Items for the US Navy General Classification Tests, Form 2, and the US Navy Tests of Reading and Arithmetical Reasoning, Form 2,* G. A. Satter, OSRD 3756, CEEB, June 8, 1944. APP-412-M3

18. *A Statistical Evaluation of the United States Navy Officer Qualification Test, Forms 2 and 3,* OSRD 3978, CEEB, Aug. 4, 1944. APP-319-M6

19. *Characteristics and Uses of Item-Analysis Data,* Herbert S. Conrad, OSRD 4034, CEEB, Aug. 19, 1944. APP-412-M4

20. *The Preparation of Norms for the Fleet Edition of the General Classification Test,* D. A. Peterson, OSRD 4242B, CEEB, Oct. 10, 1944. APP-414-M1

21. *The Construction and Validation of an Arithmetical Computation Test,* G. A. Satter and Norman Frederiksen, OSRD 4556, CEEB, Jan. 8, 1945. APP-411-M2

22. *A Statistical Evaluation of the Basic Classification Test Battery, Form 1,* Herbert S. Conrad, OSRD 4636, CEEB, May 14, 1945, pp. 77-84. APP-411-M3

23. *A Further Study of the Validity of the Arithmetical Computation Test,* Norman Frederiksen, OSRD 5302, CEEB, July 3, 1945. APP-411-M4

24. *Achievement Examination for Signalman School,* M. Duane Bown, Thelburn L. Engle, and Harold P. Bechtoldt, OSRD 5460, CEEB, Aug. 20, 1945.

25. *Summary Report on Research and Development of the Navy's Aptitude Testing Program,* Herbert S. Conrad, OSRD 6110, CEEB, Oct. 31, 1945. APP-322-M6

26. "The Relation of Item Difficulty and Inter-item Correlation to Test Variance and Reliability," Harold Gulliksen, *Psychometrika,* Vol. 10, 1945, pp. 79-91.

27. *Program Outlined for Peacetime Enlisted Classification,* Enlisted Classification Bulletin, September 1945.

CHAPTER 15

1. *Naval Aptitude Tests: A. Reliability; B. Scoring Accuracy; C. Scoring Procedures,* OSRD 1127, CEEB, Dec. 10, 1942. APP-322-M1

2. *Selection of Test Items by Correlation with an External Criterion, as Applied to the Mechanical Comprehension Test—OQT-O-2,* Harold Gulliksen, OSRD 3187, CEEB, Jan. 8, 1944. APP-325-M2

3. *The Construction and Validation of an Arithmetical Computation Test,* G. A. Satter and Norman Frederiksen, OSRD 4556, CEEB, Jan. 8, 1945. APP-411-M2

4. *A Statistical Evaluation of the Basic Classification Test Battery, Form 1,* Herbert S. Conrad, OSRD 4636, CEEB, May 14, 1945. APP-411-M3

5. *A Further Study of the Validity of the Arithmetical Computation Test,* Norman Frederiksen, OSRD 5302, CEEB, July 3, 1945. APP-411-M4

6. "Item Selection by Means of a Maximizing Function," Paul Horst, *Psychometrika,* Vol. 1, December 1936, pp. 229-244.

OSRD APPOINTEES

Applied Psychology Panel

Chiefs

W. S. Hunter
C. W. Bray

Technical Aides

C. W. Bray
J. L. Kennedy
Dael Wolfle

Members

G. K. Bennett
C. W. Bray
Leonard Carmichael
C. H. Graham
W. S. Hunter
J. M. Stalnaker
M. S. Viteles
Dael Wolfle

[redacted]

CONTRACT NUMBERS, CONTRACTORS, AND SUBJECTS OF CONTRACTS

Contract Number	*Contractor*	*Subject*
OEMsr-581	The Trustees of Tufts College Medford, Massachusetts	Studies of operator performance and sources of error in the operation of field artillery, tank destroyer, and tank guns, and in all types of antiaircraft equipment.
OEMsr-614	National Academy of Sciences Washington, D. C.	To establish a Committee to carry on research on selection, training, and related problems of manpower use; to study the design of equipment in terms of human capacities; to advise NDRC as to additional research required in the above fields and to recommend to NDRC contractors for carrying on the work.
OEMsr-700	The Trustees of the University of Pennsylvania Philadelphia, Pennsylvania	Studies in the selection and training of naval gun crews and personnel for destroyers and destroyer escorts; evaluation of methods of training free (aerial) gunners.
OEMsr-705	College Entrance Examination Board Princeton, New Jersey	Studies and experimental investigations necessary to develop the Navy's aptitude and achievement testing program.
OEMsr-815	Brown University Providence, Rhode Island	Studies and experimental investigations in connection with the selection and training of heightfinder and rangefinder operators and fire controlmen; preparation of operating instructions for Navy gun directors; improvement in the design of fire control equipment.
OEMsr-830	The Psychological Corporation New York, New York	Studies in the selection of men for communication by voice and by radio code; development of methods of training voice and radio code communication personnel; development of a device for transferring Morse code signals to typescript.
OEMsr-834	Brown University Providence, Rhode Island	Studies of methods of identifying emotionally unstable men prior to their assignment to military duty; studies of the usefulness of Battle Noise Equipment in selecting and training military personnel and retraining psychiatric casualties.
OEMsr-919	Yerkes Laboratories of Primate Biology, Inc. Orange Park, Florida	Studies of the selection and training of radar operators and the operation of radar equipment.
OEMsr-1136	Princeton University Princeton, New Jersey	Studies of the selection and training of night lookouts and of night lookout performance.
OEMsr-1171	The Regents of the University of Wisconsin Madison, Wisconsin	Studies to develop military requirements for flexible gunnery equipment which should be determined by the characteristics of Army gunners; to assist in the development of valid training methods for flexible gunners; to develop adequate methods for the selection and training of personnel for duty in gun directors.
OEMsr-1213	The President and Fellows of Harvard College Cambridge, Massachusetts	Studies of methods of testing the relative strength of various interests by determining activity preferences.
OEMsr-1298	The Trustees of Pennsylvania State College State College, Pennsylvania	Studies of job analysis, qualification, and placement of personnel in the amphibious forces.
OEMsr-1340	The Board of Trustees of the Leland Stanford Junior University Stanford University, California	Studies of methods and devices to aid in the classification and placement of men in naval jobs.
OEMsr-1372	The Trustees of the University of Southern California Los Angeles, California	Studies in the selection and training of hatchmen and winchmen specialist teams on AKA and APA vessels.

SERVICE PROJECT NUMBERS

The projects listed below were transmitted to the Office of the Executive Secretary, OSRD, from the War or Navy Department through either the War Department Liaison Officer for NDRC or the Office of Research and Inventions (formerly the Coordinator of Research and Development), Navy Department.

Service Project Number	*Title*
	Army Projects
AC-92	Investigation of the Ability of Gunners to Learn the Use of Remote Fire Control Systems of Flexible Gunnery.
AC-94	Psychological Factors in the Operation of Flexible Gunnery Equipment.
SC-67	Training Program in Voice Communication.
SC-70	Selection and Training of Radar Operators.
SC-88	Methods of Training Radio Code Operators.
SOS-6	Study of Operator Performance on All Types of Antiaircraft Equipment.
SOS-7	Development of an Activity Preference and Interest Inventory.
SOS-11	Psychological Factors in Operation and Design of Field Artillery, Tank Destroyer, and Tank Sighting Equipment.
......	Selection and Training of Height Finder Operators.
	Navy Projects
N-100	Committee on Applied Psychology in the War and Committee on Service Personnel—Selection and Training.
N-104	Determination of Reliability, Objectivity, Validity, and Independence of Medical Tests.
N-105	Selection and Training of Naval Gun Crews.
N-106	Research and Development of the Navy's Aptitude Testing Program.
N-107	Selection and Training of Radio Code Operators.
N-109	Selection and Training of Personnel Using Voice Communication Systems.
N-111	Psychological Problems in Operation of Antiaircraft Lead Computing Sights and Directors.
N-112	Study and Evaluation of Sighting Methods of Instruction Used in U. S. Naval Free Gunnery Training.
N-113	Research on a Personal Inventory and Other Tests for Selection for Special Service Tasks.
N-114	Selection and Training of Rangefinder and Radar Operators.
N-115	Selection and Training of Night Lookouts.
N-116	Selection and Training of Personnel.
N-117	Job Analysis, Qualification, and Placement of Personnel in the Amphibious Forces.
NR-106	Selection and Training of Personnel Assigned to Destroyers and Destroyer Escorts.
NS-146	Selection and Training of Radar Operators.
NS-366	Development of Morse Code Actuated Printer.

Appendix

[redacted]

UNITED STATES NAVY
OFFICER QUALIFICATION TEST

Read this first: This test is designed to enable you to provide further evidence regarding your abilities. There is no passing or failing score for the test and no one will be disqualified because of the test score alone. Your score will simply be regarded as further information about you to be added to your application record. Work as well as you can, wasting no time. No one is expected to answer all of the questions correctly. As a matter of fact, you may answer a substantial number of items incorrectly and still secure a score that is at least average.

Directions for Taking the Test

First, print your name and the other required information on the separate answer sheet furnished with this booklet. Do this now. After filling in these blanks on the answer sheet, read the rest of these directions.

The test consists of three parts. The total time for all three parts is one hour. You may divide your time as you wish, but the following division is recommended as being to your advantage:

Part I	15 minutes
Part II	15 minutes
Part III	25 minutes

This schedule leaves 5 minutes to spend on any of the parts which you have not finished, or to check any answers about which you are not sure. Remember, you have one hour to read the directions carefully and to answer as many questions as you can.

Directions and sample questions precede the questions for each part. Indicate all your answers on the separate answer sheet. Give only one answer to each question; double answers are graded as incorrect.

Do not spend too much time on any one question or on any one part. If you are uncertain, mark the answer you think is most likely to be correct.

If you have a question now or at any time during the test, raise your hand and someone will come to you.

Make no marks in this test booklet. Indicate all answers on the separate answer sheet.

Do not spend too much time on any one question or any one part.

Now read the directions for Part I on the next page and start to work at once. As soon as you finish any one page or any one part, go right on to the next without waiting for further instructions.

DO NOT STOP. GO ON TO PART I.

[redacted]

Suggested time:
15 minutes

PART I

O P P O S I T E S

Directions

Each question in this part consists of a CAPITALIZED word followed by five other words numbered from 1 to 5. Decide which one of the five numbered words is most nearly <u>opposite</u> in meaning to the capitalized word. Indicate your choice by making a heavy black mark between the lines under the corresponding number on the separate answer sheet.

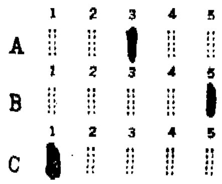
The questions below have already been marked correctly on the Sample Answer Blank. Study these questions and be sure that you understand how the answers are indicated.

A. HOT
   1 plain   2 heavy   3 cold   4 soft   5 clean

B. LAUGH
   1 giggle   2 amuse   3 snort   4 grieve   5 cry

C. LUXURY
   1 simplicity   2 delicacy   3 flourish   4 dirt   5 elegance

<u>Sample Answer Blank</u>



Now begin working on OPPOSITES. Mark each answer in the proper space on the separate answer sheet. Mark only ONE answer to each question. Remember you are to choose the word most nearly <u>opposite</u> in meaning to the CAPITALIZED word. Do not make any marks in this booklet. Work as fast and as accurately as you can.

When you finish Part I, go right on to Part II.

DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]

OPPOSITES

Mark only on separate answer sheet.

1. CREDITABLE
   1 accidental 2 unworthy 3 invisible 4 menacing 5 vigilant
2. DISTINCT
   1 infirm 2 indivisible 3 vague 4 complex 5 erroneous
3. COGNIZANT
   1 uncalled-for 2 deep-seated 3 inadequate 4 original 5 unaware
4. INVALUABLE
   1 priceless 2 expensive 3 mercenary 4 violable 5 worthless
5. PERTINENT
   1 comparable 2 irrelevant 3 courteous 4 melancholy 5 yielding
6. CARDINAL
   1 exotic 2 passionate 3 nonessential 4 fraudulent 5 unavowed
7. ZEALOUS
   1 dauntless 2 fanatical 3 humane 4 listless 5 unpatriotic
8. EMIT
   1 retreat 2 rebound 3 absorb 4 atone 5 crush
9. FLUX
   1 stability 2 piety 3 severity 4 elation 5 penury
10. VERIFICATION
    1 deprivation 2 experiment 3 inquiry 4 refutation 5 repute
11. DYNAMIC
    1 defensive 2 chaotic 3 fruitless 4 apathetic 5 onerous
12. WARY
    1 coy 2 stalwart 3 rash 4 lusty 5 shrill
13. MANDATORY
    1 unpretentious 2 derogatory 3 discretionary 4 derisive 5 palatable
14. REQUISITE
    1 dispensable 2 imperative 3 perquisite 4 prerequisite 5 unrequited
15. PERMEABLE
    1 adhesive 2 impervious 3 indulgent 4 negotiable 5 solvent
16. FUNDAMENTAL
    1 voluminous 2 approximate 3 perplexing 4 superficial 5 quiescent
17. PHLEGMATIC
    1 ungainly 2 ungracious 3 excitable 4 urbane 5 wholesome
18. RECALCITRANCE
    1 cooperation 2 resonance 3 restoration 4 relaxation 5 suavity
19. ARBITRARY
    1 absurd 2 incompetent 3 productive 4 reasoned 5 unreasoned
20. DISCREPANT
    1 fortunate 2 adept 3 consistent 4 obvious 5 intrinsic
21. APPRECIABLE
    1 concise 2 deterrent 3 discordant 4 negligible 5 ungrateful
22. EXTRANEOUS
    1 extrinsic 2 feasible 3 genuine 4 integral 5 tangible
23. FACILE
    1 arduous 2 cursory 3 infamous 4 purblind 5 factitious
24. DIMINUTION
    1 exhibition 2 anticipation 3 augmentation 4 elevation 5 transmutation
25. PROPITIOUS
    1 auspicious 2 adverse 3 controversial 4 inexperienced 5 disputatious

DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]

26. ASCERTAIN
1 conjecture 2 demur 3 discern 4 dissuade 5 underestimate

27. EXPEDITIOUS
1 dilatory 2 elusive 3 habitual 4 latent 5 scrupulous

28. PROSPECTIVE
1 evasive 2 impending 3 perspective 4 detected 5 realized

29. OBVIATE
1 necessitate 2 dissimulate 3 vilify 4 diversify 5 domineer

30. ACTUATE
1 defy 2 denounce 3 inhibit 4 obliterate 5 support

31. EFFETE
1 judicious 2 robust 3 placable 4 paltry 5 wonted

32. SALUTARY
1 civilian 2 detrimental 3 brusque 4 nutritious 5 valedictory

33. AUTOMATIC
1 deliberate 2 sporadic 3 ineffectual 4 inflammable 5 intimate

34. EXPEDIENT
1 admissible 2 devious 3 leisurely 4 gradual 5 inadvisable

35. INTELLIGIBLE
1 simulated 2 inscrutable 3 indescribable 4 ingenuous 5 innocuous

36. REDUNDANT
1 exalted 2 slack 3 emollient 4 staid 5 succinct

37. EXACERBATE
1 reverberate 2 assuage 3 temporize 4 effervesce 5 coordinate

38. DUCTILE
1 supine 2 refractory 3 pliant 4 infallible 5 indefatigable

39. DESICCATED
1 protracted 2 embellished 3 affluent 4 coarsened 5 saturated

40. PIQUANT
1 precise 2 vapid 3 anomalous 4 sagacious 5 munificent

41. ENERVATE
1 intimidate 2 energize 3 horrify 4 ingratiate 5 satiate

42. IMPROVIDENT
1 grudging 2 frugal 3 opportune 4 reverent 5 utilitarian

43. INCLEMENT
1 contrite 2 conventional 3 implacable 4 lenient 5 incessant

44. DESULTORY
1 cursory 2 methodical 3 vibrant 4 arrant 5 boisterous

45. SYNOPTIC
1 indicative 2 ocular 3 discursive 4 resultant 5 undeveloped

46. PERFUNCTORY
1 ardent 2 astute 3 constructive 4 erratic 5 peerless

47. LIMPID
1 inflated 2 sonorous 3 vigorous 4 turbid 5 pellucid

48. PALPABLE
1 adamant 2 noisome 3 blatant 4 defunct 5 incorporeal

49. SEDULOUS
1 indolent 2 agnostical 3 uncouth 4 authoritative 5 irascible

50. INTRANSIGENCY
1 culpability 2 license 3 motivation 4 serenity 5 vacillation

 DO NOT STOP. GO ON TO PART II.

 [redacted]

Suggested time:
15 minutes

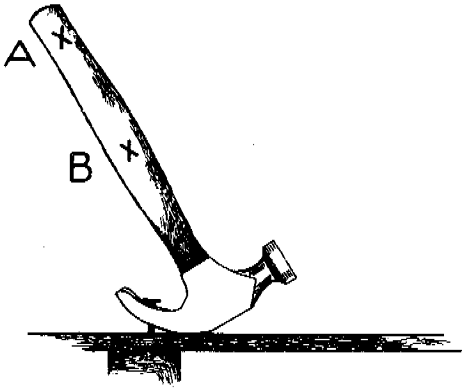
PART II

M E C H A N I C A L   C O M P R E H E N S I O N

Directions

This part consists of pictures about which questions are asked. Following each question are three answers. Decide which one of the answers is correct and make a heavy black mark between the lines under the corresponding number on the answer sheet.

The questions below have already been answered correctly on the Sample Answer Blank. Study them and be sure that you understand the problems and how the answers are indicated.



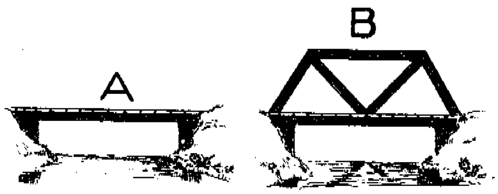
X

At which place should one grasp the handle in order to pull out the nail more easily?

(1) A

(2) B

(3) It makes no difference.



Y

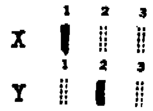
Which bridge is stronger?

(1) A

(2) B

(3) They are equally strong.

Sample Answer Blank



Now begin working on MECHANICAL COMPREHENSION. Mark each answer in the proper space on the separate answer sheet. Mark only ONE answer to each question. Do not make any marks in this booklet. Work as fast and as accurately as you can.

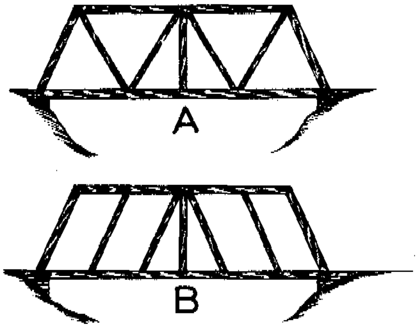
When you finish Part II, go right on to Part III.

DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]

MECHANICAL COMPREHENSION

Mark only on separate answer sheet.



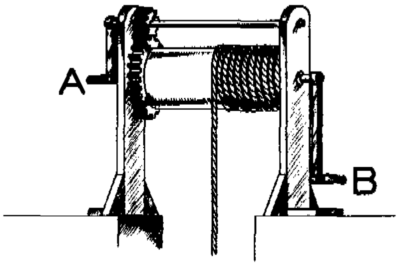
51

Which bridge is stronger?

(1) A

(2) B

(3) They are equally strong.



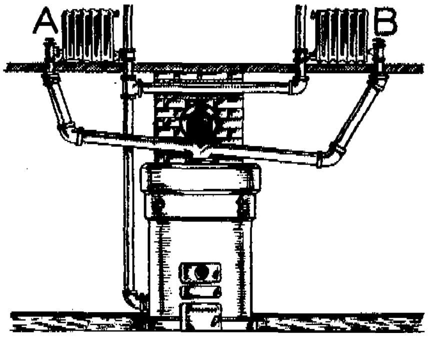
52

With which handle can one lift the heavier weight?

(1) A

(2) B

(3) It makes no difference.



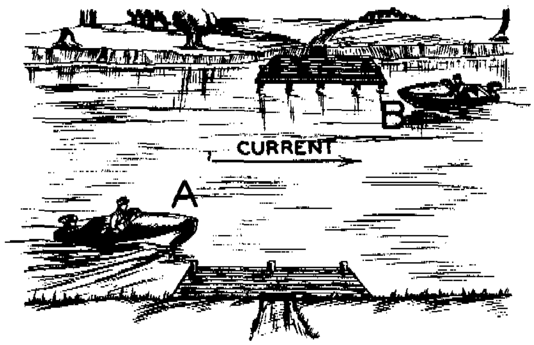
53

Which radiator will get hot first when a fire is started in the furnace?

(1) A

(2) B

(3) They will get hot at the same time.



54

Which motor boat is approaching the dock correctly for landing?

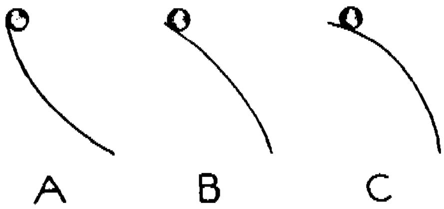
(1) A

(2) B

(3) Both are approaching correctly.

DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]



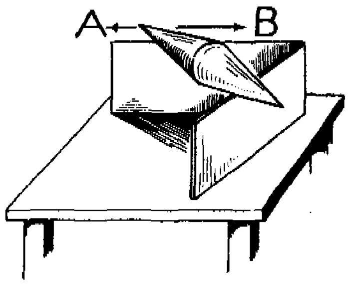
55

Which ball will roll to the bottom in the shortest time?

(1) A

(2) B

(3) C



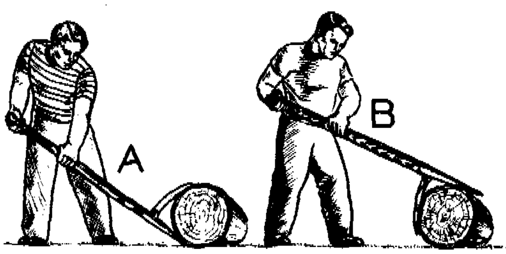
56

In which direction is the object more likely to roll?

(1) A

(2) B

(3) It is no more likely to roll in one direction than in the other.



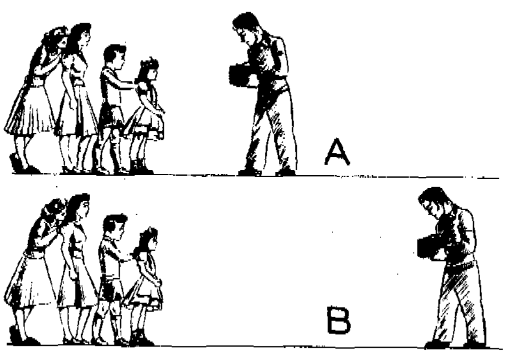
57

Which peavey is correctly placed for rolling the log to the right?

(1) A

(2) B

(3) Both are correctly placed.



58

In which position will the man get a clearer picture of both the nearest and the farthest face?

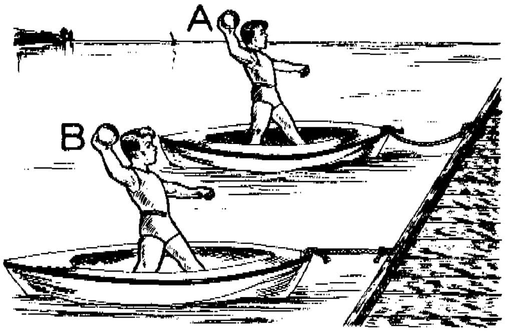
(1) A

(2) B

(3) The two faces will be equally clear in either position.

DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]



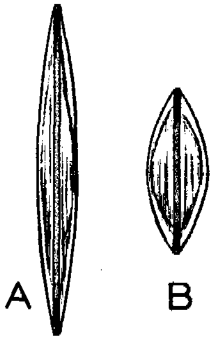
59

Which boy can throw the ball farther?

(1) A

(2) B

(3) They can throw it equally far.



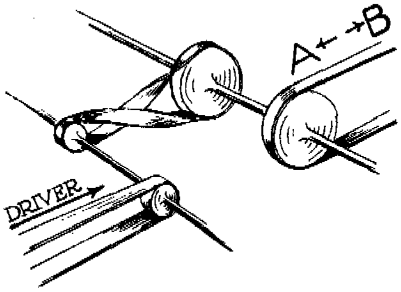
60

Which lens would be better for lighting a fire?

(1) A

(2) B

(3) They are equally good.



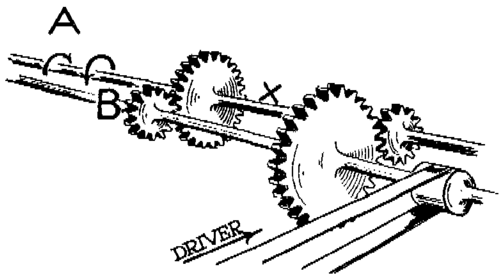
61

In which direction will the belt at the right move?

(1) A

(2) B

(3) Neither; the driver will slip.



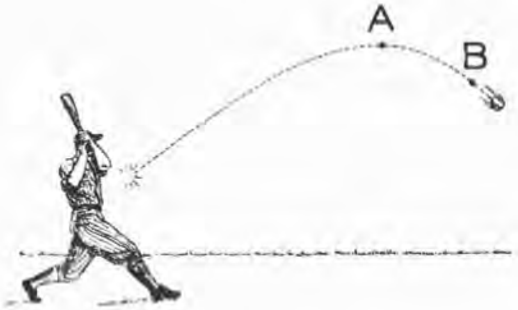
62

In which direction will shaft X turn?

(1) A

(2) B

(3) Neither; the belt will slip.



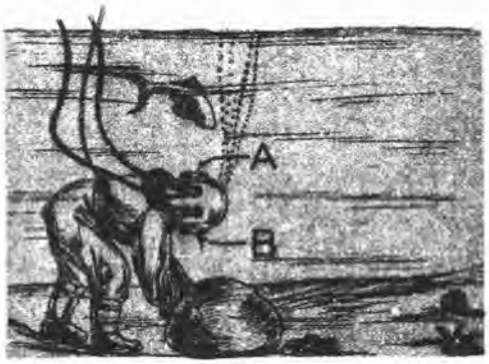
63

At which point is the baseball traveling faster?

(1) A

(2) B

(3) It travels at the same speed at both points.



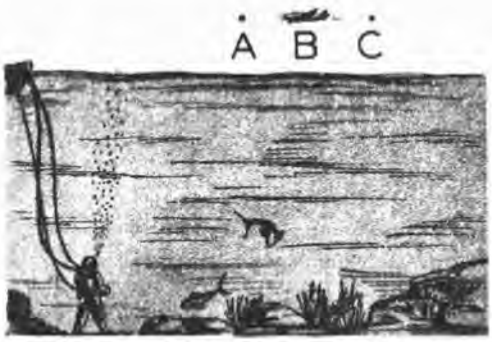
64

Against which window does the water press harder?

(1) A

(2) B

(3) It presses equally hard against both.



65

Where will the airplane appear to be to the diver?

(1) A

(2) B

(3) C

66

The bottom of the vessel of water will appear to be nearer which point?

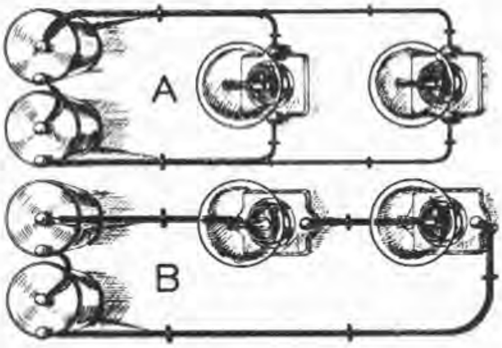
(1) A

(2) B

(3) C

DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]



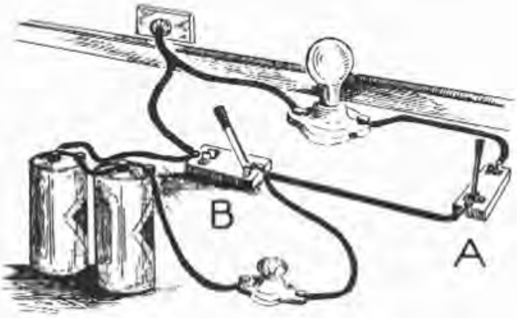
67

Which light will be brighter?

(1) A

(2) B

(3) They will be equally bright.



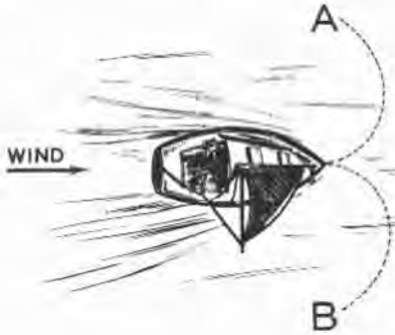
68

Which switch should be closed first?

(1) A

(2) B

(3) It makes no difference.



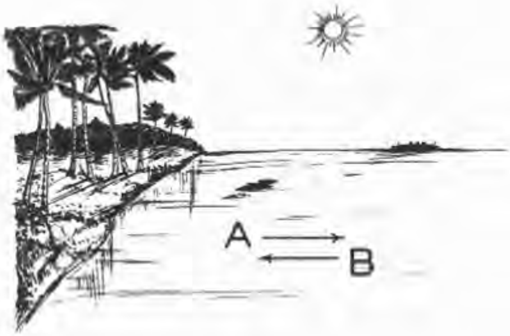
69

In which direction is it safer to turn the sailboat?

(1) A

(2) B

(3) It is equally safe to turn it in either direction.



70

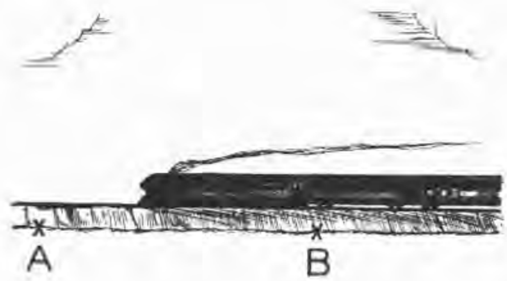
In which direction is the breeze more likely to be blowing?

(1) A

(2) B

(3) It is equally likely to be blowing in either direction.

[redacted]



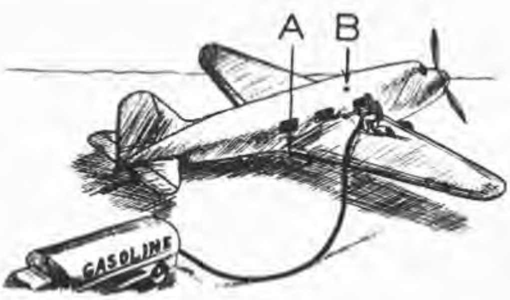
71

The train whistle will sound higher in pitch if a man is in which position?

(1) A

(2) B

(3) It will have the same pitch in either position.



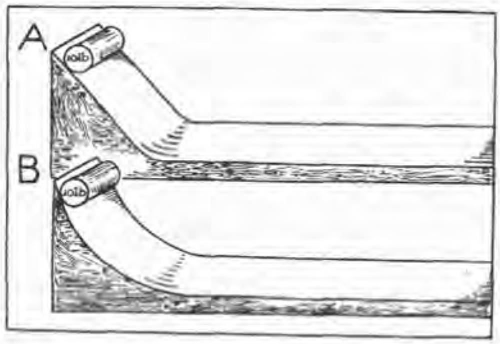
72

At which point would it be more dangerous to light a match?

(1) A

(2) B

(3) It would be equally dangerous at both points.



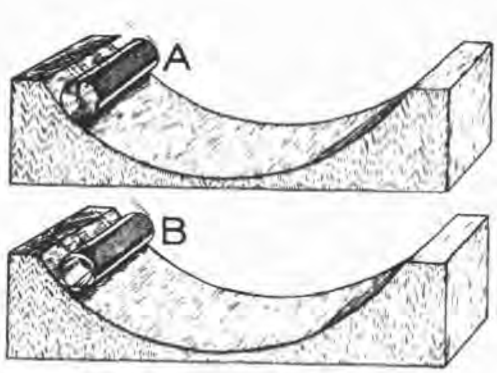
73

Which ten-pound cylinder will roll farther to the right?

(1) A

(2) B

(3) They will roll equally far.



74

Will the solid cylinder A or the hollow cylinder B roll higher up the incline at the right?

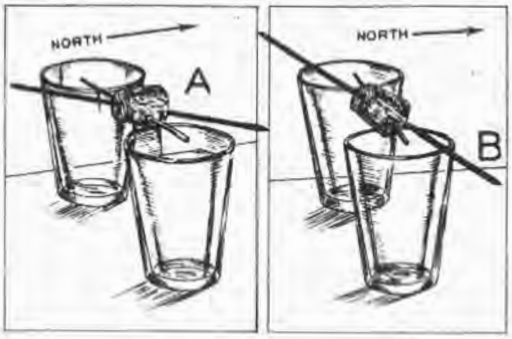
(1) A

(2) B

(3) Both will roll equally high.

DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]



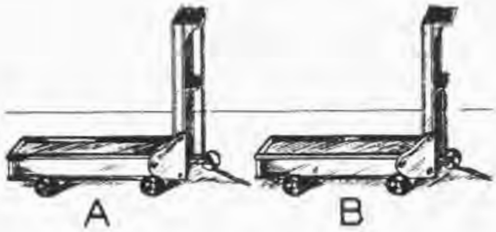
75

Which magnetized needle is farther from the squator?

(1) A

(2) B

(3) They are the same distance from the equator.



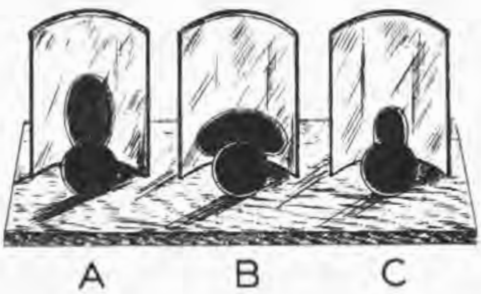
76

Which cart is more likely to tip forward when pulled by the rope?

(1) A

(2) B

(3) They are equally likely to tip.



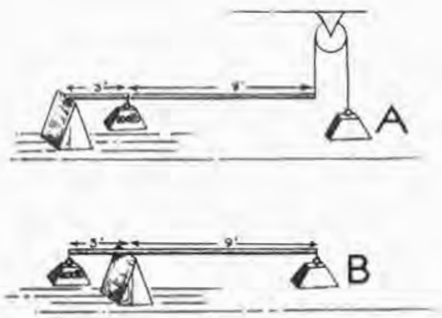
77

Which picture shows how the ball would be reflected in the curved mirror?

(1) A

(2) B

(3) C



78

Which weight would need to be heavier in order to balance the 200-pound load?

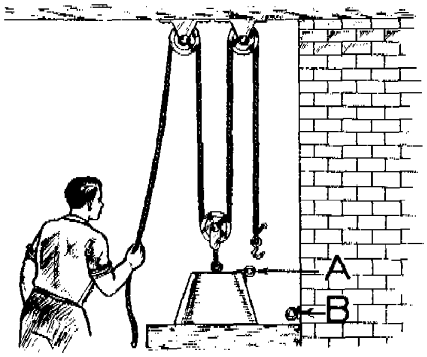
(1) A

(2) B

(3) Equal weights would be needed.

DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]



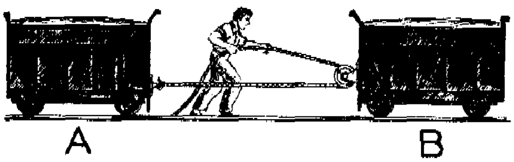
79

The weight will be easier to lift if the hook is attached to which ring?

(1) A

(2) B

(3) It will be equally easy with either ring.



80

Which car is more likely to move?

(1) A

(2) B

(3) Both are likely to move.

DO NOT STOP. GO ON TO PART III.

[redacted]

Suggested time:
25 minutes

PART III

A R I T H M E T I C A L   R E A S O N I N G

Directions

This part consists of arithmetic problems. Following each problem are five answers, lettered from (a) to (e); one of these answers is correct. Solve each problem and indicate your answer by a heavy black mark in the appropriate space on the answer sheet.

<u>Use the back of the separate answer sheet for figuring.</u>

The answers to the problems below have already been marked on the Sample Answer Blank. Solve these problems for yourself and be sure that you agree.

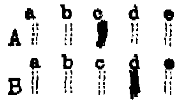
A. If three pencils cost 5 cents, how many pencils can be bought for 15 cents?

(a) 5 (b) 7 (c) 9 (d) 10 (e) 15

B. Three-fifths of the 240 uniforms for a company were delivered. How many uniforms were there yet to come?

(a) 48 (b) 60 (c) 90 (d) 96 (e) 144

<u>Sample Answer Blank</u>



Now begin working on ARITHMETICAL REASONING. Mark each answer in the proper space on the separate answer sheet. Mark only ONE answer to each problem. Do not make any marks in this booklet. Use the back of the separate answer sheet for figuring. Work as fast and as accurately as you can.

ARITHMETICAL REASONING

Mark <u>only</u> on separate answer sheet. Use <u>back</u> of the answer sheet for figuring.

81. The distance between Boston and Halifax is 450 miles. How many inches would represent this distance on a map whose scale is 50 miles to $\frac{1}{3}$ inch?

    (a) 3 (b) 4 (c) 5 (d) 9 (e) 27

82. Synthetic rubber costs 40 cents a pound whereas plantation rubber costs 15 cents a pound. The Hospital Corps uses 4,000 pounds a month in adhesive tape. How much do they save in a month on the rubber alone by using plantation rubber?

    (a) \$600 (b) \$1,000 (c) \$15,000 (d) \$60,000 (e) \$100,000

83. A man averages 90 on three arithmetic tests. His scores on the first and second tests were 85 and 94. What was his score on the third test?

    (a) 89 (b) $89\frac{1}{2}$ (c) $89\frac{2}{3}$ (d) $90\frac{1}{2}$ (e) 91

84. Two submarines start from the same base and travel in the same direction. Submarine X travels 15 miles an hour and Submarine Y travels $1\frac{1}{3}$ times as fast. How far apart are they at the end of 18 hours?

    (a) 5 mi. (b) 6 mi. (c) $67\frac{1}{2}$ mi. (d) 90 mi. (e) 630 mi.

85. A truck moves 8 feet forward with every revolution of the wheel. At 330 revolutions per minute, what is the speed of the truck in feet per second?

    (a) 5.5 (b) 41.25 (c) 44 (d) 2,475 (e) 2,640

86. A circular graph represents the classification of enlisted Naval personnel. If 225 out of each 1,000 men are apprentice seamen, how many degrees will represent this group? (A circle has 360°.)

    (a) 75° (b) 81° (c) 135° (d) 160° (e) 225°

87. Boring the 60 oval holes in a gun cooler requires 3 operations per hole. Ten round holes can be punched in a single operation. If round holes were used, how many operations would be saved on a gun cooler?

    (a) 14 (b) 170 (c) 174 (d) 186 (e) 1,800

88. A car requires 1.3 gallons of gasoline to go 20 miles at 65 miles per hour and 1 gallon to go the same distance at 25 miles per hour. How many gallons of gasoline will be saved on a trip of 50 miles by traveling at the slower speed?

    (a) $\frac{3}{4}$ (b) 1 (c) $1\frac{3}{4}$ (d) $2\frac{1}{3}$ (e) $2\frac{3}{4}$

89. A room 10 feet wide and 60 feet long is divided into six compartments, each 10 feet square. How many partitions are needed?

    (a) 5 (b) 6 (c) 7 (d) 59 (e) 60

90. A patrol boat travels due south for 12 miles, then due east for 16 miles. How many miles is it from the starting point?

    (a) 10.5 (b) 14 (c) 17 (d) 18 (e) 20

DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]

91. Six seamen can paint 4 boats in 2 days. How many days would it take two seamen to paint 16 boats if they all work at the same rate?

(a) $10\frac{2}{3}$ (b) 20 (c) $21\frac{1}{3}$ (d) 24 (e) 28

92. A reservoir is 60 feet long, 40 feet wide, and of uniform depth. How many cubic feet of water must be drawn off to lower the surface 9 inches?

(a) 75 (b) 1,600 (c) 1,800 (d) 3,200 (e) 21,600

93. An assembly process of First Aid kits requiring 300 man-hours was streamlined to require only 25 minutes. If 5,000 kits were previously assembled in 300 man-hours, how many are now assembled in an hour?

(a) $6\frac{17}{18}$ (b) 720 (c) $2,083\frac{1}{3}$ (d) 12,000 (e) 60,000

94. Two pulleys of diameter 6 inches and 10 inches are connected by a belt. How many revolutions per minute must the smaller pulley make in order that the larger pulley make 189 revolutions per minute?

(a) 68.04 (b) 113.4 (c) 315 (d) 525 (e) 1,134

95. A certain ship-building company can build a ship in 120 days. In how many days must a ship be built in order to increase the company's output by $33\frac{1}{3}$% ?

(a) 80 (b) 90 (c) 116 (d) 160 (e) 180

96. The intensity of illumination from a given light source varies inversely as the square of the distance from the source. If a certain searchlight has an intensity of 100,000 foot-candles at 10 feet, what will be the intensity of illumination in foot-candles on an airplane 3,000 feet from the source?

(a) .9 (b) 1.11 (c) 3.33 (d) 9 (e) 333.33

97. A map 6 inches long and $4\frac{1}{6}$ inches wide represents 2,500 square miles of country. On the scale to which the map is drawn an inch represents how many miles?

(a) 10 (b) 25 (c) 50 (d) 100 (e) 250

98. A cardboard container with a cover measures 8x2x3 inches. How many square inches of cardboard will be saved by making the first two measurements equal if the same capacity is retained?

(a) 0 (b) 6 (c) 12 (d) 36 (e) 48

99. Two cubes are made of the same material. The smaller cube weighs 2 pounds, and the larger weighs 128 pounds. The surface area of the larger is how many times that of the smaller?

(a) 4 (b) 8 (c) 16 (d) 64 (e) 84

100. What is the largest number of blocks each measuring 3x5x7 inches that can be packed into a box with inside dimensions of 10x14x25 inches?

(a) 18 (b) 30 (c) 32 (d) 33 (e) 34

RECHECK ANY ANSWERS YOU WISH IN THE REMAINING TIME.

 [redacted]

FORM 2

UNITED STATES NAVY

OFFICER QUALIFICATION TEST

NAVPERS 16561

Read this first: This test is designed to enable you to provide further evidence regarding your abilities. There is no passing or failing score for the test and no one will be disqualified because of the test score alone. Your score will simply be regarded as further information about you to be added to your application record. Work as well as you can, wasting no time. No one is expected to answer all of the questions correctly. As a matter of fact, you may answer a substantial number of items incorrectly and still secure a score that is at least average.

Directions for Taking the Test

This is form 2 of the test as indicated in the upper right hand corner of the booklet; so put a 2 in the space after FORM in the upper right hand corner of the separate answer sheet. Print your name and the other required information on the answer sheet. Do this now. Then read the rest of these directions.

The test consists of three parts. The total time for all three parts is one hour. You may divide your time as you wish, but the following division is recommended as being to your advantage:

Part I	15 minutes
Part II	15 minutes
Part III	25 minutes

This schedule leaves 5 minutes to spend on any of the parts which you have not finished, or to check any answers about which you are not sure. Remember, you have one hour to read the directions carefully and to answer as many questions as you can.

Directions and sample questions precede the questions for each part. Indicate all your answers on the separate answer sheet. Give only one answer to each question; double answers are graded as incorrect.

Do not spend too much time on any one question or on any one part. If you are uncertain, mark the answer you think is most likely to be correct.

If you have a question now or at any time during the test, raise your hand and someone will come to you.

Make no marks in this test booklet. Indicate all answers on the separate answer sheet.

Do not spend too much time on any one question or any one part.

Now read the directions for Part I on the next page and start to work at once. As soon as you finish any one page or any one part, go right on to the next without waiting for further instructions.

DO NOT STOP. GO ON TO PART I.

[redacted]

PART I

Suggested time:
15 minutes

O P P O S I T E S

Directions

Each question in this part consists of a CAPITALIZED word followed by five other words numbered from 1 to 5. Decide which one of the five numbered words is most nearly opposite in meaning to the capitalized word. Indicate your choice by making a heavy black mark between the lines under the corresponding number on the separate answer sheet.

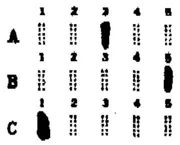
The questions below have already been marked correctly on the Sample Answer Blank. Study these questions and be sure that you understand how the answers are indicated.

A. HOT
    1 plain   2 heavy   3 cold   4 soft   5 clean

B. LAUGH
    1 giggle   2 amuse   3 snort   4 grieve   5 cry

C. LUXURY
    1 simplicity   2 delicacy   3 flourish   4 dirt   5 elegance

Sample Answer Blank



Now begin working on OPPOSITES. Mark each answer in the proper space on the separate answer sheet. Mark only ONE answer to each question. Remember you are to choose the word most nearly opposite in meaning to the CAPITALIZED word. Do not make any marks in this booklet. Work as fast and as accurately as you can.

When you finish Part I, go right on to Part II.

DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]

OPPOSITES

Mark only on separate answer sheet.

1. AUTHORIZE
   1 devaluate 2 prohibit 3 imitate 4 reprimand 5 rebel
2. MOBILE
   1 durable 2 impassive 3 fixed 4 subdued 5 tenacious
3. PERTINENT
   1 comparable 2 irrelevant 3 courteous 4 melancholy 5 yielding
4. HARASS
   1 spurn 2 loiter 3 implore 4 soothe 5 repair
5. LIABLE
   1 cautioned 2 averse 3 exempt 4 indomitable 5 unqualified
6. INVALUABLE
   1 priceless 2 expensive 3 mercenary 4 violable 5 worthless
7. COGNIZANT
   1 uncalled-for 2 deep-seated 3 inadequate 4 secret 5 unaware
8. FLUX
   1 stability 2 piety 3 severity 4 elation 5 penury
9. VERIFICATION
   1 deprivation 2 experiment 3 inquiry 4 refutation 5 repute
10. JEOPARDIZE
   1 accommodate 2 cower 3 negotiate 4 ransom 5 shield
11. EMIT
   1 retreat 2 rebound 3 absorb 4 atone 5 crush
12. INITIAL
   1 absolute 2 anterior 3 extended 4 distant 5 ultimate
13. RETRENCH
   1 revive 2 adjust 3 expand 4 display 5 reflect
14. PRETENTIOUS
   1 realistic 2 discouraged 3 convincing 4 modest 5 reliable
15. CAUSTIC
   1 bland 2 ample 3 jerky 4 eager 5 flexible
16. INTACT
   1 solaced 2 elastic 3 impaired 4 permissible 5 profound
17. FUNDAMENTAL
   1 voluminous 2 approximate 3 perplexing 4 superficial 5 quiescent
18. DISTEND
   1 heed 2 entice 3 sustain 4 vitiate 5 constrict
19. CORROBORATE
   1 disorganize 2 reduce 3 disregard 4 contradict 5 withdraw
20. PERMEABLE
   1 adhesive 2 impervious 3 indulgent 4 negotiable 5 solvent
21. VERACIOUS
   1 commendatory 2 scandalous 3 untruthful 4 negative 5 insidious
22. EBULLIENT
   1 muddy 2 epigrammatic 3 determined 4 tranquil 5 enigmatic
23. FACILE
   1 arduous 2 cursory 3 infamous 4 purblind 5 factitious
24. EXTRANEOUS
   1 extrinsic 2 feasible 3 genuine 4 integral 5 tangible
25. HACKNEYED
   1 soggy 2 reserved 3 natural 4 softened 5 original

DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]

26. PROPITIOUS
1 auspicious 2 adverse 3 controversial 4 inexperienced 5 disputatious

27. PHLEGMATIC
1 ungainly 2 ungracious 3 excitable 4 urbane 5 wholesome

28. DISSEMINATE
1 attribute 2 exhilarate 3 reassure 4 reinforce 5 suppress

29. MITIGATE
1 vindicate 2 initiate 3 aggravate 4 encourage 5 predominate

30. ACTUATE
1 defy 2 denounce 3 inhibit 4 obliterate 5 support

31. VITUPERATE
1 laud 2 constrain 3 scrutinize 4 persevere 5 purify

32. SALUTARY
1 civilian 2 detrimental 3 brusque 4 nutritious 5 valedictory

33. INCONGRUOUS
1 infinitesimal 2 enlightening 3 portentous 4 complete 5 appropriate

34. EPHEMERAL
1 enduring 2 infernal 3 immanent 4 violent 5 vulnerable

35. DESULTORY
1 apprehensive 2 methodical 3 vibrant 4 arrant 5 boisterous

36. ABET
1 substantiate 2 abhor 3 dissuade 4 err 5 abstain

37. INTRANSIGENT
1 reconcilable 2 inquisitive 3 rampant 4 sensitive 5 culpable

38. INTREPID
1 contemplative 2 staunch 3 shrewd 4 stoical 5 pusillanimous

39. PERFUNCTORY
1 enthusiastic 2 astute 3 constructive 4 erratic 5 peerless

40. DUCTILE
1 supine 2 refractory 3 pliant 4 infallible 5 indefatigable

41. ANOMALOUS
1 life-like 2 warranted 3 fantastic 4 insignificant 5 normal

42. IMPROVIDENT
1 grudging 2 frugal 3 opportune 4 reverent 5 utilitarian

43. REDUNDANT
1 exalted 2 slack 3 emollient 4 staid 5 succinct

44. SYNOPTIC
1 indicative 2 ocular 3 discursive 4 resultant 5 undeveloped

45. PERSPICACIOUS
1 devout 2 liberal 3 unimpeachable 4 supercilious 5 obtuse

46. SINUOUS
1 sudden 2 loose 3 fragile 4 straight 5 virtuous

47. TRENCHANT
1 ostensible 2 optimistic 3 unrestricted 4 polished 5 insipid

48. CATEGORICAL
1 chimerical 2 nominal 3 antithetical 4 bombastic 5 equivocal

49. INCLEMENT
1 contrite 2 conventional 3 implacable 4 lenient 5 incessant

50. INDIGENOUS
1 extant 2 wealthy 3 alien 4 fragmentary 5 straightforward

DO NOT STOP. GO ON TO PART II.

 [redacted]

Suggested time:
15 minutes

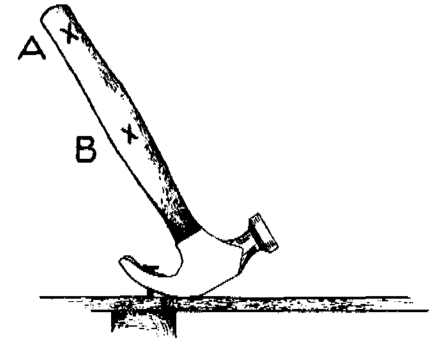
PART II

M E C H A N I C A L   C O M P R E H E N S I O N

Directions

This part consists of pictures about which questions are asked. Following each question are three answers. Decide which one of the answers is correct and make a heavy black mark between the lines under the corresponding number on the answer sheet.

The questions below have already been answered correctly on the Sample Answer Blank. Study them and be sure that you understand the problems and how the answers are indicated.



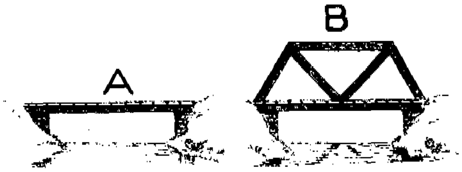
X

At which place should one grasp the handle in order to pull out the nail more easily?

(1) A

(2) B

(3) It makes no difference.



Y

Which bridge is stronger?

(1) A

(2) B

(3) They are equally strong.

Sample Answer Blank

	1	2	3
X	▮	‖	‖
Y	‖	▮	‖

Now begin working on MECHANICAL COMPREHENSION. Mark each answer in the proper space on the separate answer sheet. Mark only ONE answer to each question. Do not make any marks in this booklet. Work as fast and as accurately as you can.

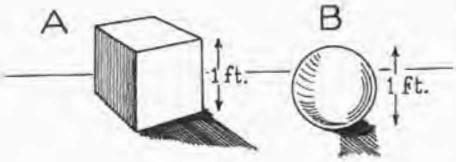
When you finish Part II, go right on to Part III.

DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]

MECHANICAL COMPREHENSION

Mark only on separate answer sheet.



51

In which case will more paint be required to cover the object?

(1) A

(2) B

(3) They will require the same amount.

52

Which camera is in the better position for taking the picture?

(1) A

(2) B

(3) The positions are equally good.

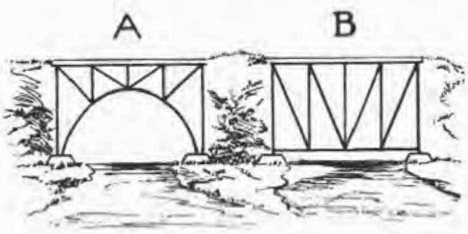
53

Which diving bell would be safer at great depths?

(1) A

(2) B

(3) One would be as safe as the other.



54

Which bridge is more likely to break in the middle?

(1) A

(2) B

(3) One is just as likely to break as the other.

DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]

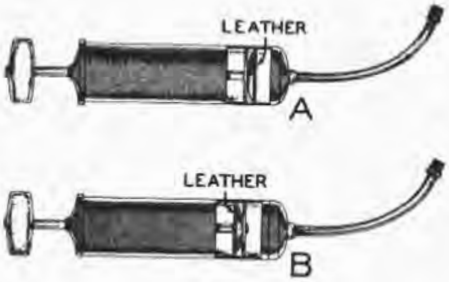
55

Which boy can throw the ball farther?

(1) A

(2) B

(3) They can throw it equally far.



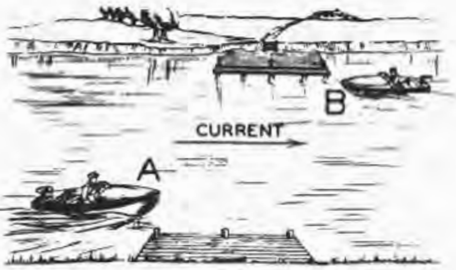
56

Which kind of pump would you use to inflate a bicycle tire?

(1) A

(2) B

(3) Either kind.



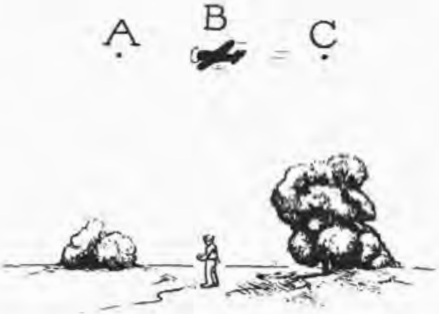
57

Which motorboat is approaching the dock correctly for landing?

(1) A

(2) B

(3) Both are approaching correctly.



58

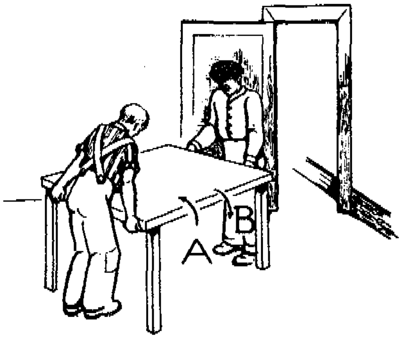
The sound of the airplane will appear to come from which point?

(1) A

(2) B

(3) C

DO NOT STOP. GO ON TO THE NEXT PAGE.



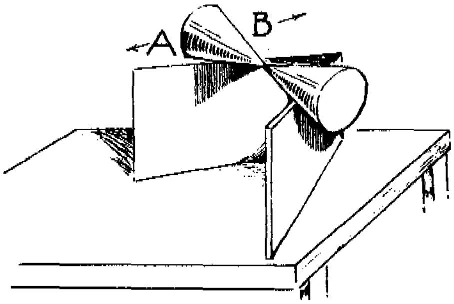
59

Which way should the men tip the table in order to get it into the next room?

(1) A

(2) B

(3) One way is as good as the other.



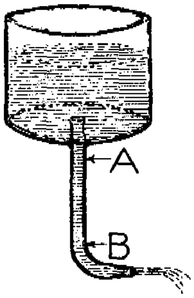
60

In which direction is the object more likely to roll?

(1) A

(2) B

(3) It is as likely to roll in one direction as in the other.



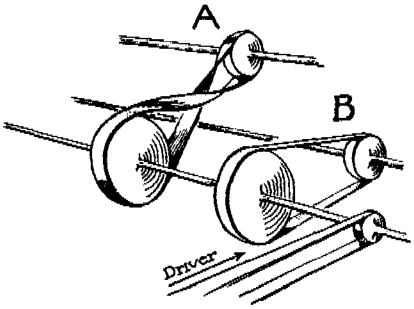
61

At which point is the water moving more rapidly?

(1) A

(2) B

(3) The speed is the same at both points.



62

Which pulley is revolving faster?

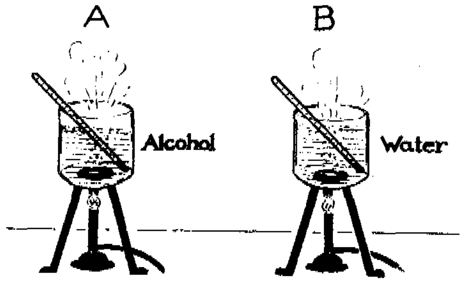
(1) A

(2) B

(3) They are revolving at about the same speed.

DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]



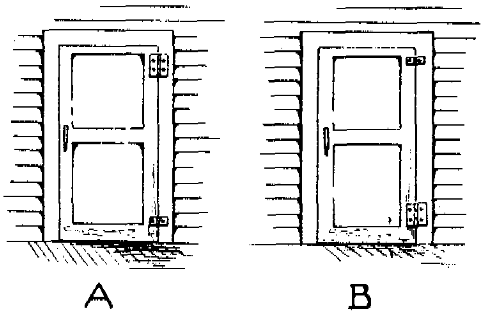
63

Which thermometer will indicate the higher temperature?

(1) A

(2) B

(3) They will show the same temperature.



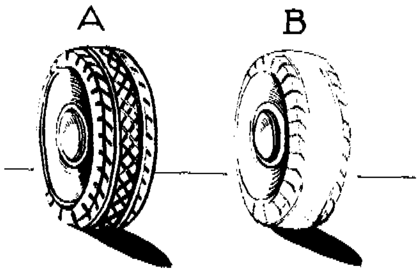
64

On which screen door is the arrangement of hinges better?

(1) A

(2) B

(3) They are equally good.



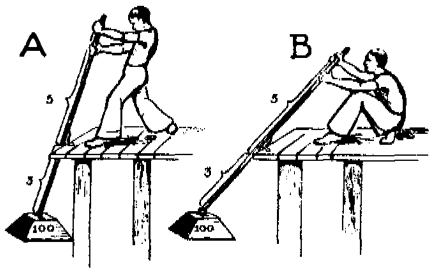
65

Which tire should be placed on the wheel that runs the speedometer in order to show higher speeds?

(1) A

(2) B

(3) It is not possible to tell.



66

Which man has to pull harder to hold the weight without moving it?

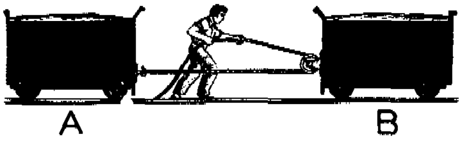
(1) A

(2) B

(3) They will have to pull equally hard.

DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]



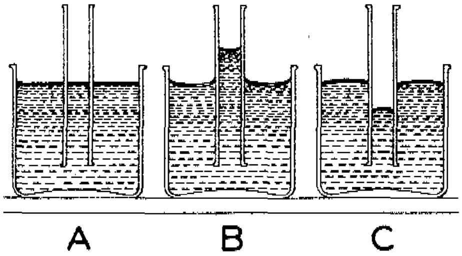
67

Which car is more likely to move?

(1) A

(2) B

(3) They are equally likely to move.



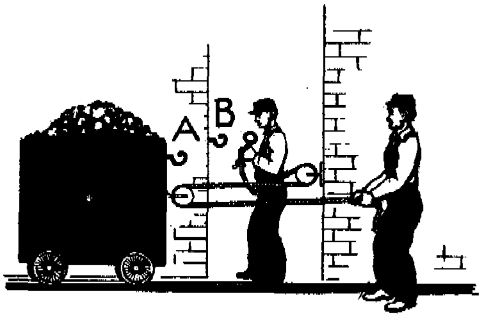
68

Which jar contains mercury?

(1) A

(2) B

(3) C



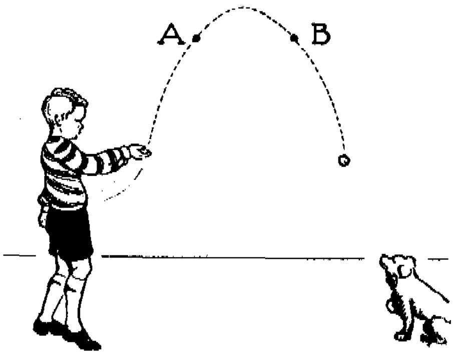
69

To which hook should the eyebolt be attached to make it easier for the man to pull the car?

(1) A

(2) B

(3) It makes no difference.



70

At which point was the stone traveling faster along its path?

(1) A

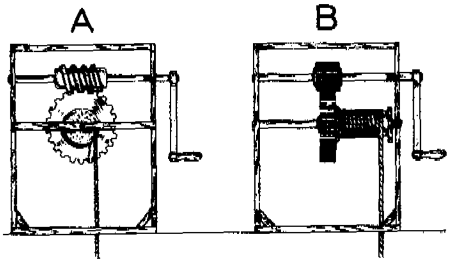
(2) B

(3) The speed was the same at both points.

DO NOT STOP. GO ON TO THE NEXT PAGE.

 [redacted]

71



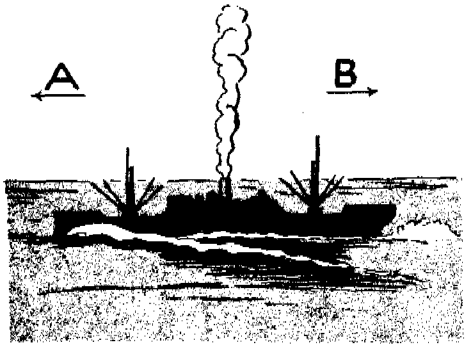
With which windlass could a man lift the heavier weight?

(1) A

(2) B

(3) There is no difference.

72



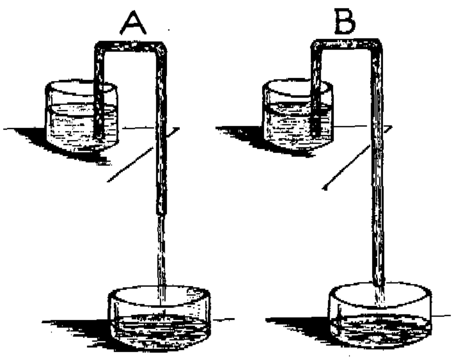
In which direction is the wind blowing?

(1) A

(2) B

(3) There is no wind.

73



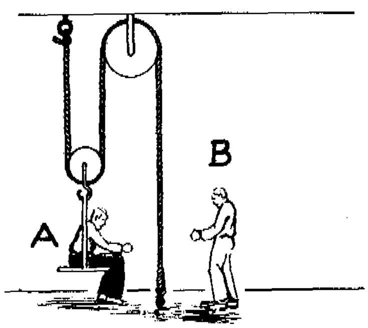
Which jar of water will empty in the shorter time?

(1) A

(2) B

(3) They will empty in the same time.

74



If A is to be raised either by pulling himself up or by having B pull him, who would have to pull the rope harder?

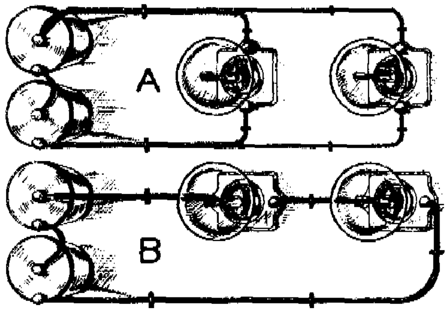
(1) A

(2) B

(3) They will have to pull equally hard.

DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]



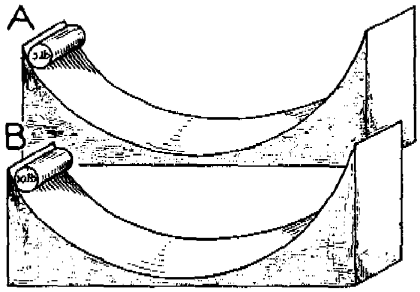
75

Which light will be brighter?

(1) A

(2) B

(3) They will be equally bright.



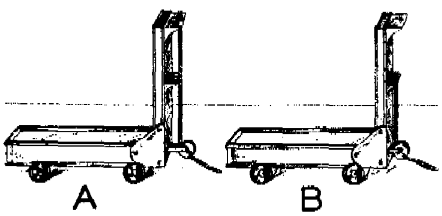
76

Which cylinder will roll higher up the incline at the right? (Disregard friction.)

(1) A

(2) B

(3) They will roll equally high.



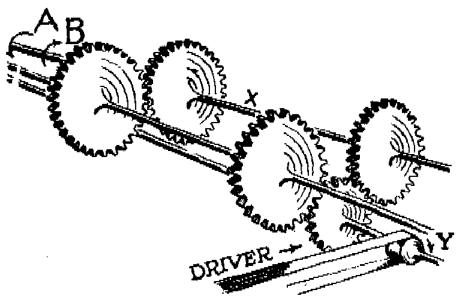
77

Which cart is more likely to tip forward when pulled by the rope?

(1) A

(2) B

(3) They are equally likely to tip.



78

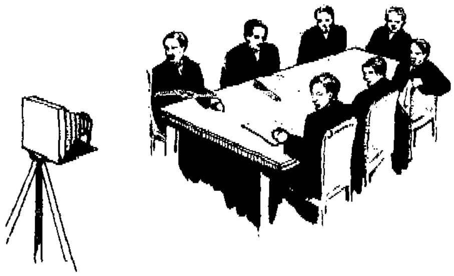
In which direction will shaft X turn, if the belt is moving in the direction shown at Y?

(1) A

(2) B

(3) Neither; the belt will slip.

[redacted]



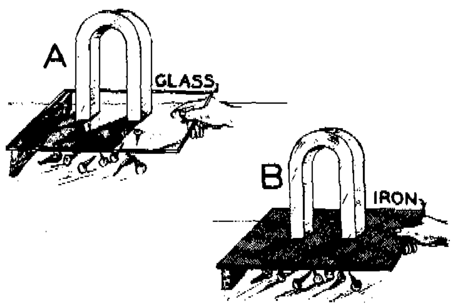
79

What size should the lens opening be in order to get a clear picture of the whole group?

(1) Large.

(2) Medium.

(3) Small.



80

Which magnet will pick up more tacks?

(1) A

(2) B

(3) They will each pick up about the same number.

DO NOT STOP. GO ON TO PART III.

[redacted]

Suggested time:
25 minutes

PART III

A R I T H M E T I C A L   R E A S O N I N G

Directions

This part consists of arithmetic problems. Following each problem are five answers, lettered from (a) to (e); one of these answers is correct. Solve each problem and indicate your answer by a heavy black mark in the appropriate space on the answer sheet.

Use the back of the separate answer sheet for figuring.

The answers to the problems below have already been marked on the Sample Answer Blank. Solve these problems for yourself and be sure that you agree.

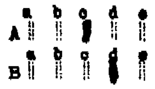
A. If three pencils cost 5 cents, how many pencils can be bought for 15 cents?

(a) 5 (b) 7 (c) 9 (d) 10 (e) 15

B. Three-fifths of the 240 uniforms for a company were delivered. How many uniforms were there yet to come?

(a) 48 (b) 60 (c) 90 (d) 96 (e) 144

Sample Answer Blank



Now begin working on ARITHMETICAL REASONING. Mark each answer in the proper space on the separate answer sheet. Mark only ONE answer to each problem. Do not make any marks in this booklet. Use the back of the separate answer sheet for figuring. Work as fast and as accurately as you can.

DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]

ARITHMETICAL REASONING

Mark only on separate answer sheet. Use back of the answer sheet for figuring.

81. In the United States Air Force a squadron consists of 12 to 15 planes. A group consists of 2 to 4 squadrons. What is the minimum number of planes that can make up a group?

    (a) 24 (b) 27 (c) 36 (d) 41 (e) 60

82. If 30% of the bombers produced in the United States are transferred to the Allies and if the Allies actually receive 90% of the bombers transferred to them, what percentage of the bombers produced in the United States are actually received by the Allies?

    (a) 3% (b) 10% (c) 27% (d) 60% (e) 120%

83. Synthetic rubber costs 40 cents a pound whereas plantation rubber costs 15 cents a pound. The Hospital Corps uses 4,000 pounds a month in adhesive tape. How much do they save in a month on the rubber alone by using plantation rubber?

    (a) $600 (b) $1,000 (c) $15,000 (d) $60,000 (e) $100,000

84. Two submarines start from the same base and travel in the same direction. Submarine X travels 15 miles an hour and Submarine Y travels $1\frac{1}{3}$ times as fast. How many miles apart are they at the end of 18 hours?

    (a) 5 (b) 6 (c) $67\frac{1}{2}$ (d) 90 (e) 630

85. There are 32 equally spaced directional points on a compass. How many degrees are there between two adjacent points? (Circle = 360°.)

    (a) $\frac{4}{45}°$ (b) $\frac{11}{120}°$ (c) $10\frac{10}{11}°$ (d) $11\frac{1}{4}°$ (e) $11\frac{19}{31}°$

86. A mixture contains by weight 1 part of cement, 3 parts of sand, and 5 parts of gravel. How many tons of gravel are needed for 27 tons of this mixture?

    (a) $\frac{5}{27}$ (b) $5\frac{2}{5}$ (c) 9 (d) 15 (e) $48\frac{3}{5}$

87. A truck moves 8 feet forward with every revolution of the wheel. At 330 revolutions per minute, what is the speed of the truck in feet per second?

    (a) 5.5 (b) 41.25 (c) 44 (d) 2,475 (e) 2,640

88. A car requires 1.3 gallons of gasoline to go 20 miles at 65 miles per hour and 1 gallon to go the same distance at 25 miles per hour. How many gallons of gasoline will be saved on a trip of 50 miles by traveling at the slower speed?

    (a) $\frac{3}{4}$ (b) 1 (c) $1\frac{3}{4}$ (d) $2\frac{1}{3}$ (e) $2\frac{3}{4}$

89. If $\frac{2}{5}$ of a crop is harvested the first day, how many more days will it take to harvest the remainder of the crop?

    (a) $\frac{3}{5}$ (b) $1\frac{1}{5}$ (c) $1\frac{1}{2}$ (d) $1\frac{2}{3}$ (e) $2\frac{1}{2}$

90. A patrol boat travels due south for 12 miles, then due east for 16 miles. How many miles is it from the starting point?

    (a) 10.5 (b) 14 (c) 17 (d) 18 (e) 20

DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]

91. The average cost of driving a new car for the first 5,000 miles is 7 cents per mile. For the first 3,000 miles the average is 8 cents. What is the average cost per mile for the remaining 2,000 miles?

(a) $5\frac{1}{2}$¢ (b) 6¢ (c) 7¢ (d) $7\frac{1}{2}$¢ (e) 25¢

92. A reservoir is 60 feet long, 40 feet wide, and of uniform depth. How many cubic feet of water must be drawn off to lower the surface 9 inches?

(a) 75 (b) 1,600 (c) 1,800 (d) 3,200 (e) 21,600

93. An assembly process of First Aid kits requiring 300 man-hours was streamlined to require only 25 minutes. If 5,000 kits were previously assembled in 300 man-hours, how many are now assembled in an hour?

(a) $6\frac{17}{18}$ (b) 720 (c) $2,083\frac{1}{3}$ (d) 12,000 (e) 60,000

94. Two pulleys having diameters of 3 inches and 5 inches are connected by a belt. How many revolutions per minute must the larger pulley make in order that the smaller make 240 revolutions per minute? (Circumference $= 2\pi r$.)

(a) 16 (b) 120 (c) 144 (d) 400 (e) 480

95. A certain trip requires 40 minutes at 45 miles per hour. How many minutes would the same trip take at 30 miles per hour?

(a) $26\frac{2}{3}$ (b) $33\frac{3}{4}$ (c) $53\frac{1}{3}$ (d) 55 (e) 60

96. A certain ship-building company can build a ship in 120 days. In how many days must a ship be built in order to increase the company's output by $33\frac{1}{3}$% ?

(a) 80 (b) 90 (c) 116 (d) 160 (e) 180

97. The daily edition of a newspaper can be printed in 60 minutes by one press and in 90 minutes by another. In how many minutes could the edition be printed if both presses were running?

(a) 30 (b) 36 (c) $37\frac{1}{2}$ (d) 75 (e) 150

98. A map 6 inches long and $4\frac{1}{6}$ inches wide represents 2,500 square miles of country. On the scale to which the map is drawn an inch represents how many miles?

(a) 10 (b) 25 (c) 50 (d) 100 (e) 250

99. The intensity of illumination from a given light source varies inversely as the square of the distance from the source. If a certain searchlight has an intensity of 100,000 foot-candles at 10 feet, what will be the intensity of illumination in foot-candles on an airplane 3,000 feet from the source?

(a) .9 (b) 1.11 (c) 3.33 (d) 9 (e) 333.33

100. Two cubes are made of the same material. The smaller cube weighs 2 pounds, and the larger weighs 128 pounds. The surface area of the larger is how many times that of the smaller?

(a) 4 (b) 8 (c) 16 (d) 64 (e) 84

RECHECK ANY ANSWERS YOU WISH IN THE REMAINING TIME.

[redacted]

FORM 3

UNITED STATES NAVY

OFFICER QUALIFICATION TEST

NAVPERS 16563

Read this first: This test is designed to enable you to provide further evidence regarding your abilities. There is no passing or failing score for the test and no one will be disqualified because of the test score alone. Your score will simply be regarded as further information about you to be added to your application record. Work as well as you can, wasting no time. No one is expected to answer all of the questions correctly. As a matter of fact, you may answer a substantial number of items incorrectly and still secure a score that is at least average.

Directions for Taking the Test

This is form 3 of the test as indicated in the upper right hand corner of the booklet; so put a 3 in the space after FORM in the upper right hand corner of the separate answer sheet. Print your name and the other required information on the answer sheet. Do this now. Then read the rest of these directions.

The test consists of three parts. The total time for all three parts is one hour. You may divide your time as you wish, but the following division is recommended as being to your advantage:

Part I	15 minutes
Part II	15 minutes
Part III	25 minutes

This schedule leaves 5 minutes to spend on any of the parts which you have not finished, or to check any answers about which you are not sure. Remember, you have one hour to read the directions carefully and to answer as many questions as you can.

Directions and sample questions precede the questions for each part. Indicate all your answers on the separate answer sheet. Give only one answer to each question; double answers are graded as incorrect.

Do not spend too much time on any one question or on any one part. If you are uncertain, mark the answer you think is most likely to be correct.

If you have a question now or at any time during the test, raise your hand and someone will come to you.

Make no marks in this test booklet. Indicate all answers on the separate answer sheet.

Do not spend too much time on any one question or any one part.

Now read the directions for Part I on the next page and start to work at once. As soon as you finish any one page or any one part, go right on to the next without waiting for further instructions.

DO NOT STOP. GO ON TO PART I.

[redacted]

PART I

Suggested time:
15 minutes

OPPOSITES

Directions

Each question in this part consists of a CAPITALIZED word followed by five other words numbered from 1 to 5. Decide which one of the five numbered words is most nearly <u>opposite</u> in meaning to the capitalized word. Indicate your choice by making a heavy black mark between the lines under the corresponding number on the separate answer sheet.

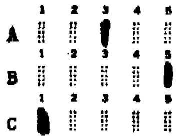
The questions below have already been marked correctly on the Sample Answer Blank. Study these questions and be sure that you understand how the answers are indicated.

A. HOT
1 plain   2 heavy   3 cold   4 soft   5 clean

B. LAUGH
1 giggle   2 amuse   3 snort   4 grieve   5 cry

C. LUXURY
1 simplicity   2 delicacy   3 flourish   4 dirt   5 elegance

<u>Sample Answer Blank</u>



Now begin working on OPPOSITES. Mark each answer in the proper space on the separate answer sheet. Mark only ONE answer to each question. Remember you are to choose the word most nearly <u>opposite</u> in meaning to the CAPITALIZED word. Do not make any marks in this booklet. Work as fast and as accurately as you can.

When you finish Part I, go right on to Part II.

DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]

OPPOSITES

Mark only on separate answer sheet.

1. ECONOMICAL
   1 plentiful 2 selfish 3 cheap 4 scarce 5 wasteful
2. DISTINCT
   1 infirm 2 indivisible 3 vague 4 complex 5 erroneous
3. ZEALOUS
   1 dauntless 2 fanatical 3 humane 4 listless 5 unpatriotic
4. FEASIBLE
   1 rigid 2 impracticable 3 insufficient 4 clandestine 5 inaccessible
5. PROLONG
   1 forestall 2 curtail 3 disdain 4 pause 5 vanish
6. CONCUR
   1 violate 2 criticize 3 dissent 4 vex 5 persist
7. INVALIDATE
   1 flatter 2 emulate 3 entrust 4 claim 5 confirm
8. CONTINGENT
   1 urgent 2 antiquated 3 essential 4 independent 5 continual
9. ACTUAL
   1 unknown 2 foreign 3 mysterious 4 imaginary 5 perceptible
10. COPIOUS
    1 internal 2 intended 3 clever 4 scanty 5 explanatory
11. SELECTIVE
    1 indiscriminate 2 careless 3 unbiased 4 deficient 5 undistinguished
12. DILATE
    1 muffle 2 refract 3 complicate 4 contract 5 animate
13. ACRID
    1 astringent 2 mild 3 refined 4 sodden 5 fluffy
14. DYNAMIC
    1 defensive 2 chaotic 3 fruitless 4 apathetic 5 onerous
15. INDUBITABLE.
    1 trifling 2 questionable 3 revocable 4 theoretical 5 indefensible
16. SOMNOLENT
    1 imperious 2 alert 3 critical 4 scurrilous 5 considerate
17. MANDATORY
    1 unpretentious 2 derogatory 3 discretionary 4 derisive 5 palatable
18. REQUISITE
    1 dispensable 2 imperative 3 perquisite 4 prerequisite 5 unrequited
19. REPROOF
    1 affection 2 laxity 3 sympathy 4 praise 5 suspicion
20. WARY
    1 coy 2 stalwart 3 rash 4 lusty 5 shrill
21. RECALCITRANCE
    1 cooperation 2 resonance 3 restoration 4 relaxation 5 suavity
22. DIMINUTION
    1 exhibition 2 anticipation 3 augmentation 4 elevation 5 transmutation
23. DISCREPANT
    1 fortunate 2 adept 3 consistent 4 obvious 5 intrinsic
24. APPRECIABLE
    1 concise 2 deterrent 3 discordant 4 negligible 5 ungrateful
25. ARBITRARY
    1 absurd 2 incompetent 3 productive 4 reasoned 5 unreasoned

DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]

26. IGNOMINY
1 elegance 2 satisfaction 3 honor 4 order 5 notoriety

27. DELETERIOUS
1 humorous 2 beneficial 3 harmonious 4 aristocratic 5 submissive

28. EXPEDITIOUS
1 dilatory 2 elusive 3 habitual 4 latent 5 scrupulous

29. EFFETE
1 judicious 2 robust 3 placable 4 paltry 5 wonted

30. ASCERTAIN
1 conjecture 2 demur 3 discern 4 dissuade 5 underestimate

31. TRUCULENT
1 overbearing 2 amiable 3 short-sighted 4 reticent 5 niggardly

32. ESTIMABLE
1 contemptible 2 unmistakable 3 inexplicable 4 fabulous 5 meager

33. VOLUPTUOUS
1 diffident 2 ascetic 3 unvaried 4 rapacious 5 fastidious

34. PROSPECTIVE
1 evasive 2 impending 3 perspective 4 detected 5 realized

35. INTELLIGIBLE
1 simulated 2 inscrutable 3 indescribable 4 ingenuous 5 innocuous

36. FATUOUS
1 embarrassed 2 groveling 3 astute 4 matchless 5 relentless

37. OBVIATE
1 necessitate 2 dissimulate 3 vilify 4 diversify 5 domineer

38. ABRIDGE
1 amend 2 amplify 3 demolish 4 improve 5 dissociate

39. SALIENT
1 implausible 2 explicit 3 paradoxical 4 imperceptible 5 sluggish

40. PIQUANT
1 precise 2 vapid 3 anomalous 4 sagacious 5 munificent

41. EXPEDIENT
1 admissible 2 devious 3 leisurely 4 gradual 5 inadvisable

42. LIMPID
1 inflated 2 sonorous 3 vigorous 4 turbid 5 pellucid

43. ATTENUATE
1 materialize 2 resound 3 intensify 4 impregnate 5 buoy

44. GERMANE
1 inapplicable 2 illogical 3 figurative 4 salubrious 5 involved

45. GLIB
1 sedate 2 tactful 3 halting 4 sound 5 rude

46. ENERVATE
1 intimidate 2 energize 3 horrify 4 ingratiate 5 satiate

47. EXACERBATE
1 reverberate 2 assuage 3 temporize 4 effervesce 5 coordinate

48. PALPABLE
1 adamant 2 noisome 3 blatant 4 defunct 5 incorporeal

49. PENCHANT
1 displeasure 2 antipathy 3 unconcern 4 exuberance 5 ineptitude

50. EFFULGENT
1 tenebrous 2 incarnate 3 gelatinous 4 adulterated 5 osseous

DO NOT STOP. GO ON TO PART II.

[redacted]

Suggested time:
15 minutes

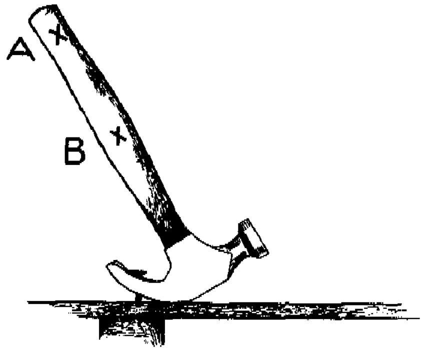
PART II

M E C H A N I C A L   C O M P R E H E N S I O N

Directions

This part consists of pictures about which questions are asked. Following each question are three answers. Decide which one of the answers is correct and make a heavy black mark between the lines under the corresponding number on the answer sheet.

The questions below have already been answered correctly on the Sample Answer Blank. Study them and be sure that you understand the problems and how the answers are indicated.



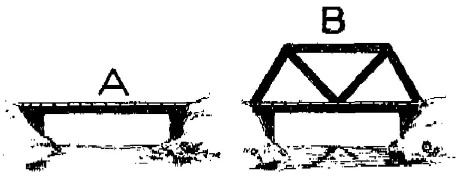
X

At which place should one grasp the handle in order to pull out the nail more easily?

(1) A

(2) B

(3) It makes no difference.



Y

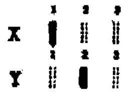
Which bridge is stronger?

(1) A

(2) B

(3) They are equally strong.

Sample Answer Blank



Now begin working on MECHANICAL COMPREHENSION. Mark each answer in the proper space on the separate answer sheet. Mark only ONE answer to each question. Do not make any marks in this booklet. Work as fast and as accurately as you can.

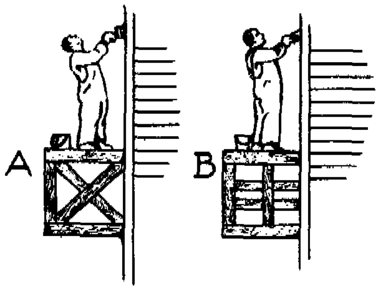
When you finish Part II, go right on to Part III.

DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]

MECHANICAL COMPREHENSION

Mark only on separate answer sheet.



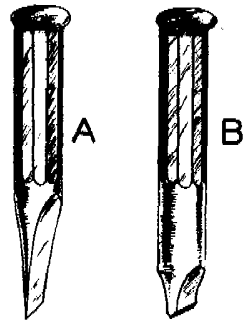
51

Which scaffold is more likely to break?

(1) A

(2) B

(3) They are equally likely to break.



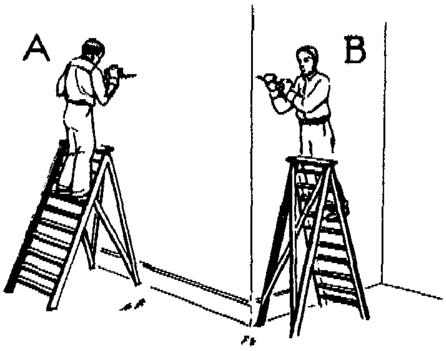
52

Which chisel is better for cutting very hard material?

(1) A

(2) B

(3) One is as good as the other.



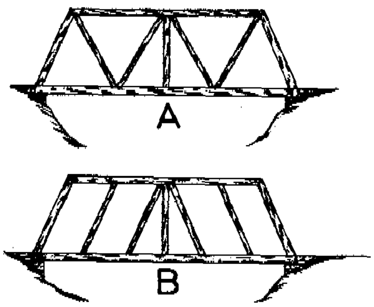
53

Which man is more likely to fall if he pushes hard on the drill?

(1) A

(2) B

(3) They are equally likely to fall.



54

Which bridge is stronger?

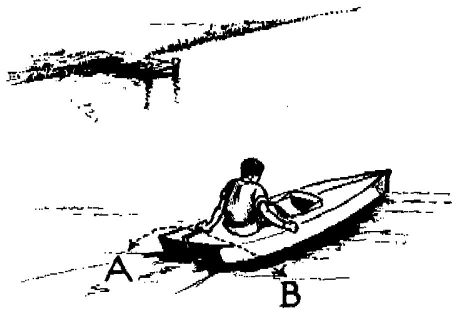
(1) A

(2) B

(3) They are equally strong.

DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]



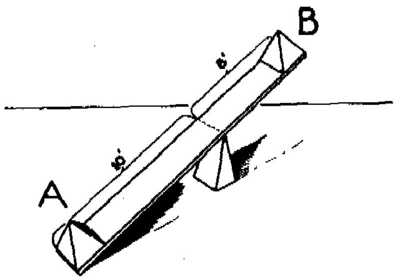
55

Which way should the tiller be pushed in order to approach the dock?

(1) A

(2) B

(3) Either way is correct.



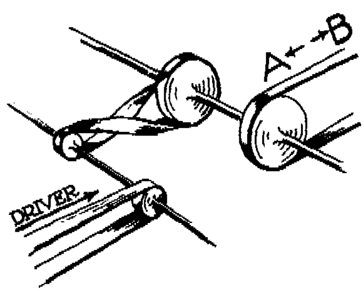
56

Which weight is heavier?

(1) A

(2) B

(3) It is impossible to say.



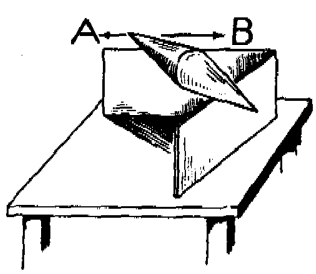
57

In which direction will the belt at the right move?

(1) A

(2) B

(3) Neither; the driving belt will slip.



58

In which direction is the object more likely to roll?

(1) A

(2) B

(3) It is as likely to roll in one direction as in the other.

DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]

59

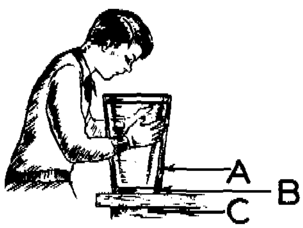
Which peavey is correctly placed for rolling the log to the right?

(1) A

(2) B

(3) Both are correctly placed.

60



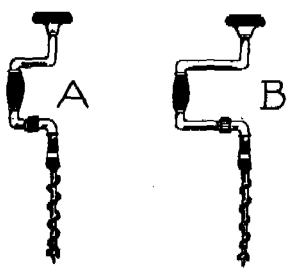
The bottom of the vessel of water will appear to be nearer which point?

(1) A

(2) B

(3) C

61



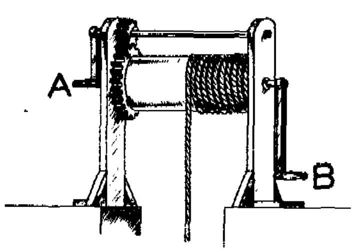
With which tool would it be easier to drill hard wood?

(1) A

(2) B

(3) There is no difference.

62



With which handle can one pull the heavier weight out of the well?

(1) A

(2) B

(3) It makes no difference.

DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]

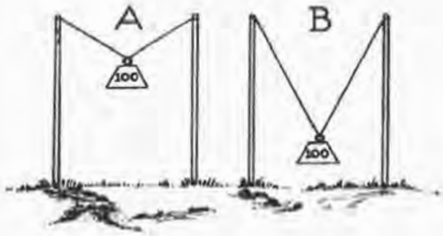
63

Against which window does the water press harder?

(1) A

(2) B

(3) It presses equally hard against both.



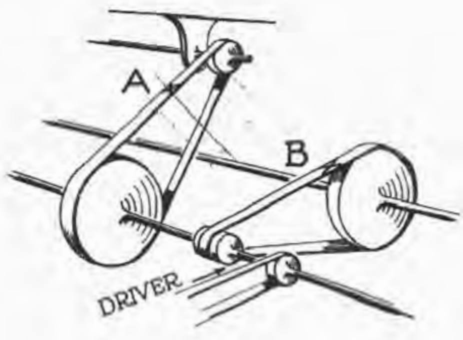
64

Which weight puts more strain on the posts?

(1) A

(2) B

(3) There is no difference.



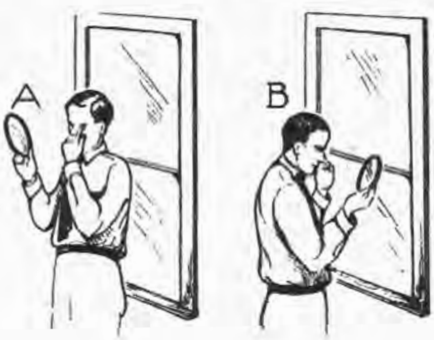
65

Which point will move faster?

(1) A

(2) B

(3) They will move at the same speed.



66

Which position gives the better lighting for getting something out of the eye?

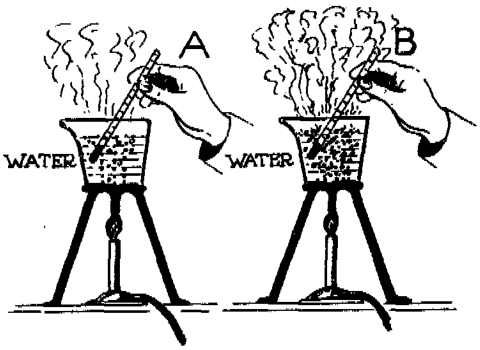
(1) A

(2) B

(3) They are equally good.

DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]



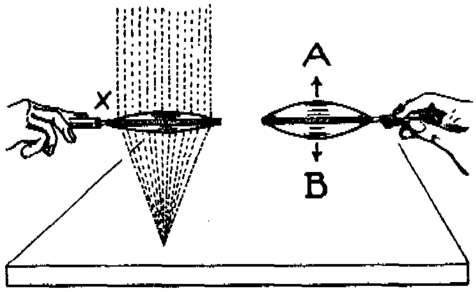
67

Which thermometer will show the higher temperature?

(1) A

(2) B

(3) They will show the same temperature.



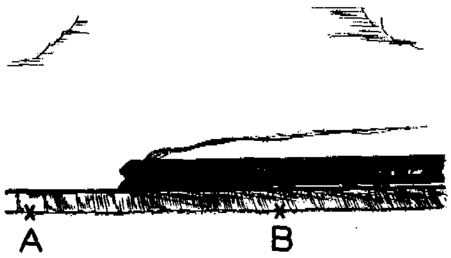
68

If lens X brings the sun's rays to a focus on the top of the table, in what direction should the other lens be moved so that it will do the same?

(1) A

(2) B

(3) It should not be moved at all.



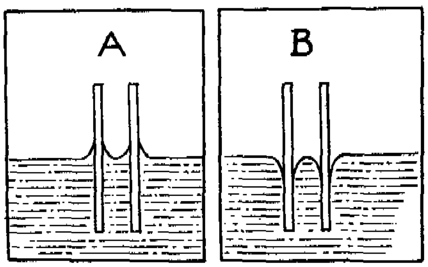
69

The train whistle will sound higher in pitch if a man is in which position?

(1) A

(2) B

(3) It will have the same pitch in either position.



70

In which tube will the liquid rise?

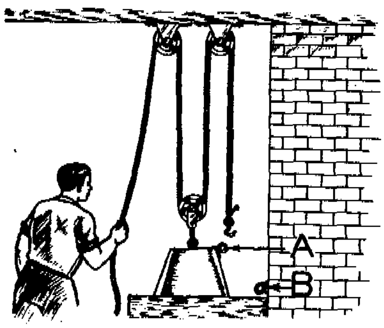
(1) A

(2) B

(3) It will rise in both tubes.

DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]



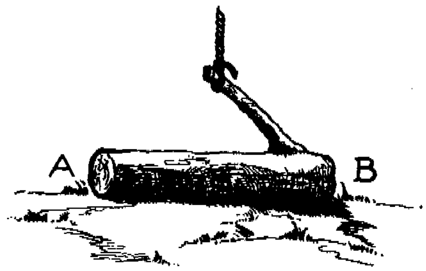
71

The weight will be easier to lift if the hook is attached to which ring?

(1) A

(2) B

(3) It will be equally easy to lift either way.



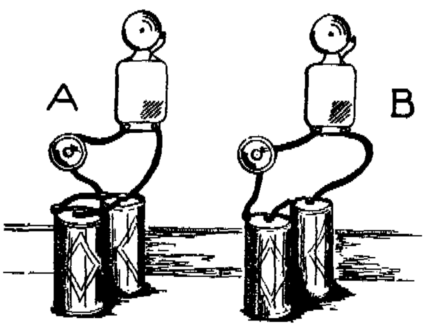
72

When the log is lifted by the rope, which end will hang down?

(1) A

(2) B

(3) Neither; the log will remain almost horizontal.



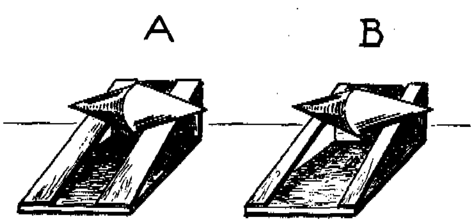
73

Which bell will ring more loudly?

(1) A

(2) B

(3) One will ring as loudly as the other.



74

Which object will roll to the bottom more quickly?

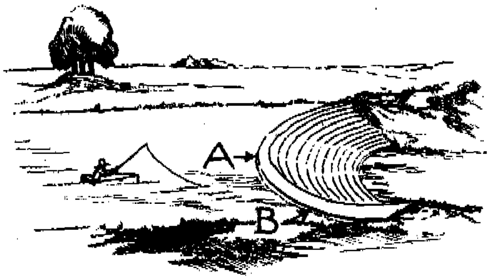
(1) A

(2) B

(3) They will reach it at the same time.

DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]



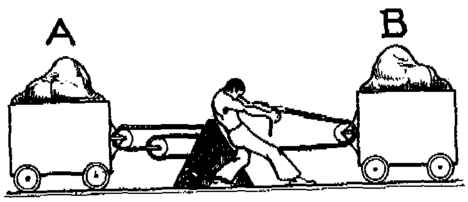
75

At which point on the dam is the water pressure greater?

(1) A

(2) B

(3) It is the same at both points.



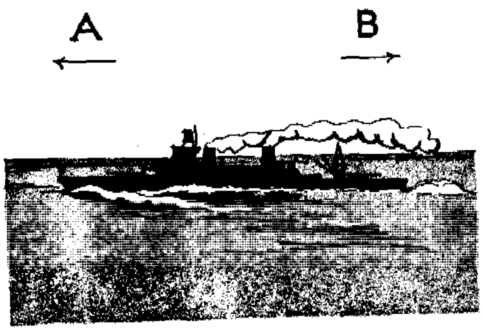
76

Which car is more likely to move?

(1) A

(2) B

(3) They are equally likely to move.



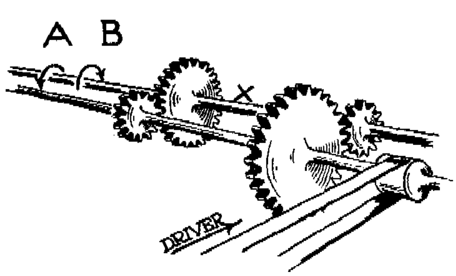
77

In which direction is the wind blowing?

(1) A

(2) B

(3) It is not possible to tell.



78

In which direction will shaft X turn?

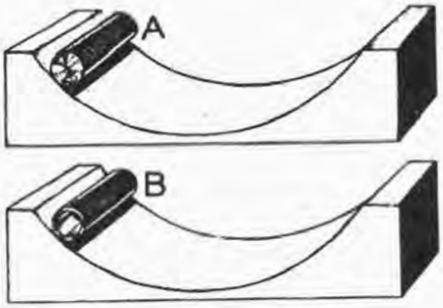
(1) A

(2) B

(3) Neither; the belt will slip.

DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]



79

Will the solid cylinder A or the hollow cylinder B roll higher up the incline at the right? (Disregard friction.)

(1) A

(2) B

(3) They will roll equally high.

80

In which case is the pressure on the man's head greater?

(1) A

(2) B

(3) The pressure is exactly the same in both cases.

DO NOT STOP. GO ON TO PART III.

[redacted]

Suggested time:
25 minutes

PART III

A R I T H M E T I C A L   R E A S O N I N G

Directions

This part consists of arithmetic problems. Following each problem are five answers, lettered from (a) to (e); one of these answers is correct. Solve each problem and indicate your answer by a heavy black mark in the appropriate space on the answer sheet.

Use the back of the separate answer sheet for figuring.

The answers to the problems below have already been marked on the Sample Answer Blank. Solve these problems for yourself and be sure that you agree.

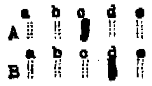
A. If three pencils cost 5 cents, how many pencils can be bought for 15 cents?

(a) 5 (b) 7 (c) 9 (d) 10 (e) 15

B. Three-fifths of the 240 uniforms for a company were delivered. How many uniforms were there yet to come?

(a) 48 (b) 60 (c) 90 (d) 96 (e) 144

Sample Answer Blank



Now begin working on ARITHMETICAL REASONING. Mark each answer in the proper space on the separate answer sheet. Mark only ONE answer to each problem. Do not make any marks in this booklet. Use the back of the separate answer sheet for figuring. Work as fast and as accurately as you can.

DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]

ARITHMETICAL REASONING

Mark only on separate answer sheet. Use back of the answer sheet for figuring.

81. If the passenger rate for air transportation is reduced from 5 cents a mile to 3 cents a mile, how many dollars will be saved on a 1,000 mile trip?

    (a) \$20 (b) \$30 (c) \$40 (d) \$80 (e) \$200

82. The distance between Boston and Halifax is 450 miles. How many inches would represent this distance on a map whose scale is 50 miles to $\frac{1}{3}$ inch?

    (a) 3 (b) 4 (c) 5 (d) 9 (e) 27

83. In estimating the student-capacity of a class room, 18 square feet of floor space are allowed per student. How many students can be accommodated in a room 36 feet long and 27 feet wide?

    (a) $3\frac{1}{2}$ (b) 24 (c) 45 (d) 54 (e) 17,496

84. A man averages 90 on three arithmetic tests. His scores on the first and second tests were 85 and 94. What was his score on the third test?

    (a) 89 (b) $89\frac{1}{2}$ (c) $89\frac{2}{3}$ (d) $90\frac{1}{2}$ (e) 91

85. In general an invading force should be 50% larger than the defending force to achieve minimum superiority. If the Allies attempt an invasion of Nazi territory, how many divisions should the Allies have to achieve minimum superiority over 30 Nazi divisions?

    (a) 20 (b) 45 (c) 50 (d) 60 (e) 75

86. A circular graph represents the classification of enlisted Naval personnel. If 225 out of each 1,000 men are apprentice seamen, how many degrees will represent this group? (Circle = 360°.)

    (a) 75° (b) 81° (c) 135° (d) 160° (e) 225°

87. A room 10 feet wide and 60 feet long is divided into six square compartments, each 10 feet on a side. How many partitions are needed?

    (a) 5 (b) 6 (c) 7 (d) 59 (e) 60

88. The approximate weight (in pounds) of a shell is one-half the cube of the diameter (in inches) of the gun. What is the approximate weight in pounds of a shell used in a gun 4 inches in diameter?

    (a) 2 (b) 6 (c) 8 (d) 16 (e) 32

89. Step A in manufacturing a certain product can be done by one person at the rate of 25 units every hour. Step B can be done by one person at the rate of 45 units every 2 hours. How many workers will be required at step A to keep 10 workers busy at step B?

    (a) 9 (b) $9\frac{2}{9}$ (c) 11 (d) $11\frac{1}{9}$ (e) 18

90. A boat has a speed of 12 miles per hour in still water. How many hours would it take to go 12 miles down stream when the current is 4 miles per hour?

    (a) $\frac{2}{3}$ (b) $\frac{3}{4}$ (c) $1\frac{1}{3}$ (d) $1\frac{1}{2}$ (e) 3

DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]

91. Six seamen can paint 4 boats in 2 days. How many days would it take two seamen to paint 16 boats if they all work at the same rate?

(a) $10\frac{2}{3}$ (b) 20 (c) $21\frac{1}{3}$ (d) 24 (e) 28

92. Two pulleys of diameter 6 inches and 10 inches are connected by a belt. How many revolutions per minute must the smaller pulley make in order that the larger pulley make 189 revolutions per minute?

(a) 68.04 (b) 113.4 (c) 315 (d) 525 (e) 1,134

93. A patrol boat, maintaining a constant speed, went due east for an hour and then due north for one hour and twenty minutes. If it then steered a straight return to its base, how many hours was the return trip?

(a) $\sqrt{\frac{7}{3}}$ (b) $\frac{1}{3}\sqrt{7}$ (c) $\frac{\sqrt{61}}{5}$ (d) $1\frac{2}{3}$ (e) $2\frac{1}{3}$

94. In 1942 the railroads carried $\frac{1}{3}$ more freight than they carried in 1941 but used only $\frac{2}{3}$ as many cars. The average load per car in 1942 was how many times as great as that for 1941?

(a) $\frac{2}{9}$ (b) $\frac{1}{2}$ (c) $\frac{8}{9}$ (d) 1 (e) 2

95. Because of the additional power furnished by a new high-octane gasoline, bombers will be able to increase their bomb loads 3 tons for every 10 tons of bombs now carried. At this rate, how many tons of bombs will a Flying Fortress be able to carry if it can now carry 3 tons of bombs?

(a) .9 (b) 3.3 (c) 3.9 (d) 6.33 (e) 10

96. If an American dollar is worth $1.10 in Canadian money, $100.00 in Canadian money is worth how much in American money (to the nearest cent)?

(a) $90.00 (b) $90.91 (c) $99.90 (d) $101.00 (e) $110.00

97. A bin 12 feet long, 10 feet wide, and 8 feet high is filled with grain. How many feet must the surface be lowered if 360 bushels are to be removed? (.8 bu. = 1 cu. ft.)

(a) $2\frac{2}{5}$ (b) 3 (c) $3\frac{3}{4}$ (d) $4\frac{1}{4}$ (e) $4\frac{3}{5}$

98. A bomber can be assembled in 20 days. If the output is to be increased 25%, in how many days must a bomber be assembled?

(a) 15 (b) 16 (c) 18 (d) 25 (e) $26\frac{2}{3}$

99. Two cubes are to be made of the same material and so constructed that the surface of the larger is 36 times that of the smaller. If the smaller weighs 3 pounds, how many pounds does the larger weigh?

(a) 12 (b) 18 (c) 108 (d) 216 (e) 648

100. If 5,000 workers produce 75 pounds of drugs per day, what percentage of the total output is produced by one worker?

(a) .015% (b) .02% (c) .67% (d) 1.5% (e) 2%

RECHECK ANY ANSWERS YOU WISH IN THE REMAINING TIME.

[redacted]

PERSONAL INVENTORY, LONG FORM

Format B

DO NOT MARK ON THIS BOOKLET

In this questionnaire you are to give information which will help others understand you. You are to indicate certain things about your job preferences, interests, etc.

In each question you will always have two answers to choose between -- the one on the left side of the page, and the one on the right. Choose the answer which fits you best. Even if neither answer fits you very well, you must choose the one that fits you better than the other. Mark your choice on the Answer Sheet, and not on the test booklet.

The examiner will show you how to fit the Answer Sheet along side the booklet so that the number of the question you are working on is right next to the same number on the Answer Sheet. Simply indicate your answer by filling in the proper space on the answer sheet. If for a given question you choose the answer on the left, then fill in the left-hand space, or if you choose the answer on the right, fill in the right-hand space on the Answer Sheet. Do this by making a heavy mark which completely fills the space between the two dotted lines. If you change an answer, be sure to erase completely.

Remember, you must always choose one answer for each question, but never both. Be sure not to skip any questions.

Prepared by Brown University
Project N-113
Applied Psychology Panel
National Defense Research Committee

[redacted]

	L	R	
1	I prefer a job that challenges my ability	I prefer a job I can do without any difficulty	1
2	I like to take a chance	I like to think things through before I do them	2
3	I've "got guts"	I can be depended on	3
4	I like to have people do things my way	I like to have people figure things out for me	4
5	The few jobs I've had I've been very much interested in	I've worked at all kinds of things	5
6	I haven't spent much time on a farm	Most of my life has been spent on a farm	6
7	I usually plan things ahead and then keep plugging away	I find that things will usually work themselves out	7
8	I prefer working my way up in one job	When I change jobs I like to go to another town	8
9	I got my first job near my home town Yes	.... No	9
10	I left jobs because I didn't have to work Yes	.... No	10
11	School didn't bother me any more than the next guy	I left school because I had enough of it	11
12	In school the other boys went their way and I went mine	I was happiest when I was with the gang	12
13	In civilian life I would rather be a mechanic	In civilian life I would rather be a florist	13
14	I like office work	I like heavier work	14
15	I like working by myself	I like working with others	15
16	Since leaving school I've averaged less than one job a year (or I haven't been out of school a year)	Since leaving school I have averaged more than one job a year	16
17	I'd rather work than take it easy	Only a good job is worth keeping	17
18	I prefer interesting work	I prefer a good boss	18
19	I could work better if I could get more sleep	The average amount of sleep is enough for me	19
20	My employers were more than fair in their opinions of me	I think I was a better worker than my employers thought	20
21	I have been out of a job for three months or more in the last five years (or since leaving school) Yes	.... No	21
22	I take life easy	I tend to worry	22
23	Being bossed around gripes me	A fellow needs a pretty firm boss	23
24	People will step all over you if you aren't careful	Most people will give you a break	24
25	Most of my friends are luckier than I am	I have my share of lucky breaks	25

 [redacted]

	L	R	
26	I wish I didn't have the blues so often	I wish I could be more serious about things	26
27	I miss the comforts I used to have	I miss my friends	27
28	I am somewhat disappointed in my family	I am proud of my family	28
29	I make friends less easily than the average fellow	I make friends more easily than the average fellow	29
30	I left school because I preferred working to studying Yes	.... No	30
31	I seek excitement	I avoid excitement	31
32	When I'm excited or upset I feel sweaty	When I'm excited or upset I don't show it much	32
33	Once I've made a decision I figure "That's that," and forget it	Often after I make a decision I wonder if I've done the right thing	33
34	If I go drinking I can get it out of my system in an evening	Once I've started drinking I enjoy going on a real "bat" and getting it out of my system	34
35	When I get excited I find it hard to talk straight	When I get excited I talk better than I usually do	35
36	After exertion I feel hungry	After exertion I feel dizzy	36
37	I pay people back	I let people get away with things	37
38	I'm very careful to take medicine whenever I need it	Most people aren't as sick as they think they are	38
39	I like most any kind of food	I have a poor appetite	39
40	I've got a steady girl or I'm married	I haven't settled down to one girl yet	40
41	I get embarrassed easily	I seldom get embarrassed	41
42	I think I might like to watch a surgical operation sometime	The sight of blood upsets me	42
43	I like to listen to the radio	I prefer a bang-up party	43
44	I want to be different from my dad	I admire my dad's (or guardian's) way of doing things	44
45	I usually need a drink to get along in a social situation Yes	.... No	45
46	The boys know better than to trifle with me	I am careful not to get in trouble with the boys	46
47	I have more headaches than the average person Yes	.... No	47
48	I like to stay put	I've gone "on the bum"	48
49	I feel nauseated more after eating	I feel nauseated more after being excited	49
50	I usually like to go to bed	Sometimes I dread going to bed	50

[redacted]

	L	R	
51	When excited I feel weak	When excited I feel stronger	51
52	I sometimes faint for no good reason Yes	.... No	52
53	Being up in high places never bothers me	I feel uncomfortable about being in high places	53
54	I can't think straight when I'm mad	I think better when I'm mad	54
55	I'd like a chance to think without being disturbed	I prefer being with the boys	55
56	Drinking alone is usually a waste of good liquor Yes	.... No	56
57	I think I could have done better in school if I hadn't changed schools so often	I think I could have done better in school if I had studied more	57
58	I drink now and then	I'm proud to say I've never touched a drop	58
59	I sometimes drink because it helps me to forget Yes	.... No	59
60	I can easily handle a pint of hard liquor without getting dead drunk Yes	.... No	60
61	I wish I wouldn't feel so tired	I wish I could have a more responsible job	61
62	I wish I had more responsibility	I wish I had more self-confidence	62
63	I wish I weren't so nervous	I wish I wouldn't keep putting things off	63
64	I wish people would stop trying to get my goat	I wish I weren't so affected by what people say	64
65	I wish I could get myself to take more chances	I wish worrying wouldn't make me sick to my stomach	65
66	I wish I could have more excitement	I wish I weren't bothered by bad dreams	66
67	I wish I weren't so definite in my opinions	I wish people wouldn't talk me into doing things	67
68	I wish I weren't so different from other people	I wish I weren't so hasty	68
69	I wish I didn't have so many aches and pains	I wish I wouldn't keep changing my mind	69
70	I wish my feelings weren't so easily hurt	I wish I had more time to spend with my friends	70
71	I would like to be a baseball player	I would like to be an artist	71
72	I would rather be an auto racer	I would rather be a poet	72
73	I wish I wouldn't have so many ups and downs	I wish I could make myself talk more	73
74	I've always been somewhat uneasy	Nothing ever bothers me	74
75	Somehow I never could find enough to do in my free time	My free time always seemed to be filled	75

[redacted]

	L	R	
76	I think being afraid is worse	I think lying is worse	76
77	I think cops usually do as much harm as good Yes	.... No	77
78	I sometimes get confused without any apparent reason Yes	.... No	78
79	I am reckless	I am calm	79
80	You have to worry more about heart trouble than most people think Yes	.... No	80
81	I wish people would mind their own business more	I wish I could make myself talk more	81
82	I am more nervous	I am more easy going	82
83	I am more cocky	I am more quiet	83
84	I wish I could have more responsibility	I wish I wouldn't worry so much	84
85	As a kid I sometimes ran away from home	I liked to stay around home too much when I was a kid	85
86	As a boy I usually went to my friends' houses to play	As a boy I usually brought my friends to my house to play	86
87	In school things came easy to me	In school I had to work for what I got	87
88	In school I had a pretty good idea of what I wanted	A lot of things taught in school seemed like a waste of time	88
89	As a child I was often punished when I didn't deserve it	When I was a child I deserved whatever punishment I got	89
90	My family is somewhat disappointed in me	My family is fairly pleased with me	90
91	As a boy I had my share of cuts and bruises	I had my share of dizzy spells as a boy	91
92	As a kid I played with the older boys	As a kid I played with the younger boys	92
93	I didn't let anybody in school put anything over on me	In school the boys were usually square with me	93
94	As a child I had nightmares Yes	.... No	94
95	In school I played hookey more than the average fellow	In school I played hookey less than the average fellow	95
96	I was a sickly child	I was an active child	96
97	I didn't waste my time hanging around after school	I spent quite a bit of time in school activities	97
98	I have never had a head injury	I have had a head injury	98
99	In school I often got in trouble with my teachers	In school I was a little afraid of my teachers	99
100	In school I liked to be by myself	In school I preferred to be with a gang	100

[redacted]

	L		R		
101	I left school because I needed the money	Yes	....	No	101
102	I stopped school before I got to high school	Yes	....	No	102
103	I graduated from high school	Yes	....	No	103
104	I got my first steady job (or entered the armed services) within 4 months after leaving school	Yes	....	No	104
105	I repeated a grade in school	Yes	....	No	105
106	I have felt bad more from staying up late at night		I have felt bad more from riding cars, trains, or busses		106
107	My heart sometimes speeds up for no reason at all	Yes	....	No	107
108	I've had my share of backaches, headaches, and stomach aches	Yes	....	No	108
109	Being in a warm room makes me drowsy		Being in a warm room makes me dizzy		109
110	I have never gone to a doctor for headaches or dizzy spells		I have occasionally gone to a doctor for headaches or dizzy spells		110
111	I have been annoyed more by sore throat		I have been annoyed more by constipation and loose bowels		111
112	I have felt bad more often from colds		I have felt bad more often from nausea		112
113	I have sometimes wet the bed (urinated) since the age of 10		I have never wet the bed since the age of 10		113
114	If I were a research doctor, I would rather find a cure for cancer		If I were a research doctor, I would rather find a cure for epilepsy		114
115	I had an unhappy childhood		My childhood was happy		115
116	My own parents brought me up	Yes	....	No	116
117	My family went its own way pretty much		My family took part in a lot of things around town		117
118	My parents are separated	Yes	....	No	118
119	I was a sensitive kid		I was a happy-go-lucky kid		119
120	My parents were away from home a good deal when I was a boy	Yes	....	No	120
121	I've had my share of sickness		I've had my share of happiness		121
122	Before entering the armed forces I went to a doctor or a hospital for my nerves		Before entering the armed forces I never went to a doctor or a hospital for my nerves		122
123	I have felt bad more from head cold		I have felt bad more from dizziness		123

[redacted]

	L	R	
124	I've always had a good idea of what I'm going to do	Military service seems like a good solution to my problems	124
125	I joined the CCC's more than once Yes	.... No	125
126	People talk to each other too much about things that concern me	People are frank and open with me	126
127	Having people kid me makes me mad	Bumping into something makes me mad	127
128	A friend is a person who helps you get ahead	A friend is a person who understands you	128
129	Has anyone in your family ever gone to a doctor or hospital for nervousness? Yes	.... No	129
130	My family was on relief for a while Yes	.... No	130
131	My family was always quarreling	My family stuck too close together	131
132	Our family scraps often came after someone had been drinking	Drinking never was the cause of our family scraps	132
133	My father worked for more than 8 years on one job or for one company Yes	.... No	133
134	I have been divorced Yes	.... No	134
135	I am married or intend to get married sometime Yes	.... No	135
136	Something makes me say and do things I don't mean	I can get along in most any situation	136
137	The hours at night seem long Yes	.... No	137
138	Being in a small closed-in place never bothers me	Being in a small closed-in place makes me feel uneasy	138
139	Did you ever wake up and find your tongue bitten? Yes	.... No	139
140	Have you bitten your nails since you were 15? Yes	.... No	140
141	Have you ever had a convulsion? Yes	.... No	141
142	Have you ever been arrested for anything other than a traffic violation? Yes	.... No	142
143	Have you ever been sent to jail or reform school? Yes	.... No	143
144	Have you ever been a patient in a mental hospital? Yes	.... No	144
145	Do you think fighting men should get pensions after the war? Yes	.... No	145

[redacted]

Part B

OFFICERS'

PERSONAL INVENTORY

Format 4

Do not write or mark on this booklet. You are to indicate your answers on a separately provided Answer Sheet.

In each question you will have two answers to choose between—the one on the left side of the page, and the one on the right. Choose the answer which fits you best, and mark your Answer Sheet accordingly; that is, if you choose the answer on the left, fill in the L space on the Answer Sheet, or if you choose the answer on the right, fill in the R space. Always fill in either one or the other, but never both. Even if neither answer fits you very well, you must choose the one that fits you better than the other.

[redacted]

	L		R	
1	I was an only child	Yes	.... No	1
2	As a boy I went to camp		I never went to a boys' camp	2
3	I was a Sea Scout or First Class Boy Scout	Yes	.... No	3
4	As a boy I knew Morse Code		As a boy I did not know Morse Code	4
5	I couldn't swim when I was 10		I learned to swim before I was 10	5
6	I was a sickly child		I was an active child	6
7	I didn't waste my time hanging around after school		I spent quite a bit of time in school activities	7
8	I was a dare-devil when I was a boy		I was overly cautious when I was a boy	8
9	I have grown more tolerant as I have grown older		I have grown more cautious as I have grown older	9
10	As a child I had nightmares	Yes	.... No	10
11	I was happiest when I was with the gang		In school the other boys went their way and I went mine	11
12	As boys we were always thinking up new initiations and hazings		We didn't go in for initiations very much	12
13	As a boy I was ringleader of the gang		A friend of mine led the gang	13
14	I was a happy-go-lucky kid		I was a sensitive kid	14
15	In our neighborhood games one of us was always getting pretty well banged up		I was never injured in any game as a boy	15
16	In school I spent less than half of my free time in athletics		In school I spent about half of my free time in athletics	16
17	I have coached, or helped coach, the other fellows on the team		I've never coached, or helped coach, a team	17
18	I consider myself a good athlete		I consider myself about an average athlete	18
19	I prefer an opponent of equal skill		I prefer playing against an opponent much better than myself	19
20	I enjoy the tired feeling that comes after strenuous exercise		I don't believe in exercising too strenuously	20
21	After exertion I feel dizzy		After exertion I feel hungry	21
22	I usually eat a good meal		I usually eat very little	22
23	I enjoy camping out and roughing it		I've never cared much for camping	23
24	At an amusement park I prefer the penny arcade		At an amusement park I prefer the roller-coaster	24
25	Most problems are like difficult games		Problems and games have little similarity	25
26	Work well done is sufficient reward in itself		Work well done should be praised	26
27	I prefer a job that challenges my ability		I prefer a job I can do without any difficulty	27
28	I wish I wouldn't feel so tired		I wish I could have a more responsible job	28
29	If I have an unpleasant job to do, I try to get it over with as quickly as possible		If I have an unpleasant job to do, I like to take my time about doing it	29

[redacted]

	L	R	
30	I find any job tiring after a while	I have endurance to resist fatigue and nervous strain	30
31	In civilian life I would rather be a florist	In civilian life I would rather be a mechanic	31
32	The sight of blood upsets me	I think I might like to watch a surgical operation sometime	32
33	I prefer to be a passenger in an automobile	I prefer to be the driver of an automobile	33
34	I have many practical skills	I'm better at handling ideas	34
35	I am very much interested in a few things	I am interested in all kinds of things	35
36	I prefer working under definite orders	I like to be on my own	36
37	I've never been particularly mechanically inclined	I've always had a mechanical bent	37
38	When things go wrong I call in the repair man	I am an expert on fixing things	38
39	I don't seem to be very inventive	I have a couple of schemes that I've always felt might make good inventions	39
40	Some of the boys are much better at the job than I am	I can do my work as well as anybody	40
41	The average amount of sleep is enough for me	I could work better if I could get more sleep	41
42	I never stayed away from work unless I was sick	I often had other reasons for staying away from work	42
43	I wish I had more self-confidence	I wish I had more responsibility	43
44	I wish I wouldn't keep putting things off	I wish I weren't so nervous	44
45	I don't like to attempt a job unless I'm pretty sure of the outcome	I'd tackle anything within reason	45
46	When my plans fall through, I usually have another "trick up my sleeve"	When my plans fall through, I am usually at a loss as to what to do next	46
47	I make it a point to know my stuff thoroughly	I make it a point to know the essentials, but to save a good part of my energy for emergencies	47
48	My ideas are usually unique	My ideas are usually practical	48
49	I'm inclined to look at the practical side of things	I'm inclined to look at the humorous side of things	49
50	I work best under pressure	I work worst under pressure	50
51	I like working with others	I like working by myself	51
52	I wish I wouldn't worry so much	I wish I could have more responsibilty	52
53	I never doubt my ability to get things done	I never doubt my ability to do things myself	53
54	I prefer to do individual work	I prefer to control or handle people	54
55	I find it easier to explain things orally	I find it easier to explain things in writing	55
56	I get disgusted when people don't catch on	I get annoyed when people don't catch on	56
57	I usually ask people to do things	I usually tell people to do things	57
58	I am more interested in things	I am more interested in people	58

 [redacted]

	L	R	
59	I'm more interested in what people are thinking	I'm more interested in what people are doing	59
60	People are frank and open with me	People talk to each other too much about things that concern me	60
61	Our telephone operators at home often gave me the wrong numbers	I never had any trouble giving information over the telephone	61
62	I prefer to talk to a stranger by phone	I prefer to talk to a stranger in person	62
63	I often keep people waiting.	I'm always on time for appointments	63
64	My free time always seemed to be filled	Somehow I never could find enough to do in my free time	64
65	I pick my friends carefully	I like to meet new people	65
66	I haven't settled down to one girl yet	I've got a steady girl or I'm married	66
67	Most people are o. k. when you get to know them	There is always somebody in every crowd that nobody can get along with	67
68	I let people solve their own problems	I'm often asked to solve my friends' problems	68
69	I often make and break little promises	A promise should never be broken under any circumstances	69
70	I remember people's names easily	Ordinarily I don't pay much attention to people's names	70
71	The second time I meet a person I call him by his first name	I don't call people by their first names unless I know them well	71
72	I would rather have athletic friends	I would rather have less active friends	72
73	I write letters mostly to friends	I write letters mostly to relatives	73
74	I'm always one of the group	Sometimes I feel apart or aloof from the others	74
75	I'm too hard to please	I'm too easy to please	75
76	I tend to worry	I take life easy	76
77	I avoid excitement	I seek excitement	77
78	When excited I feel stronger	When excited I feel weak	78
79	I feel nervous more after excitement is over	I feel nervous more during stress and excitement	79
80	I can act calm and undisturbed when I am afraid	I can't appear very self-assured when things are uncertain	80
81	I'm looking forward to a crisis to show my stuff	I'm not sure just how I would act in a crisis	81
82	I never let myself get in a tight spot	When in a tight spot I can always find a way out	82
83	I think better in an emergency	I think better when I have time to reflect	83
84	I rarely regret my speedy decisions	I regret many of my speedy decisions	84
85	I prefer to make my own decisions	I usually seek reliable advice before making decisions	85
86	I've always been a deep thinker	I've always had good judgment	86
87	I can make important decisions immediately	I prefer to think things over before deciding	87
88	I usually keep my ideas to myself	I like to discuss my ideas with others	88

[redacted]

	L	R	
89	A good argument provokes thought	I dislike any argument	89
90	I prefer to discuss theoretical matters	I prefer to discuss practical things	90
91	To avoid friction, I accept decisions with which I don't agree	I try to change people around to my point of view	91
92	When a party gets dull I liven it up	I prefer to let others start the ball rolling	92
93	I have never led any group	I am respected among my friends as a leader	93
94	I can explain things clearly	I have a forceful manner	94
95	I'm a hustler	I'm a diplomat	95
96	I'm seldom phased	Sometimes I get mixed up	96
97	I'm somewhat slow and deliberate in my actions	I'm energetic in my actions	97
98	I am more nervous	I am more easy going	98
99	I usually show my emotions	I usually hide my emotions	99
100	I am calm	I am reckless	100
101	Sometimes I dread going to bed	I usually like to go to bed	101
102	Our neighbors at home were too noisy	Our neighbors at home were too sedate	102
103	When I once get mad, it takes me a while to get over it	I flare up easily, but get over it quickly	103
104	I sometimes get confused without any apparent reason No	.... Yes	104
105	I think better when I'm mad	I can't think straight when I'm mad	105
106	When I get excited I talk better than I usually do	When I get excited I find it hard to talk straight	106
107	I do a certain amount of good wholesome griping	I tend to keep my grievances to myself	107
108	I think more quickly than the average person	I think about as quickly as the average person	108
109	I enjoy finding my way around in a strange city	I prefer being in a city where I know my way around	109
110	I wish worrying wouldn't make me sick to my stomach	I wish I could get myself to take more chances	110
111	I wish I had more time to spend with my friends	I wish my feelings weren't so easily hurt	111
112	People consider me a clear thinker	People consider me a good mixer	112
113	People consider me ambitious	People consider me unselfish	113
114	I can get along in most any situation	Something makes me say and do things I don't mean	114
115	I tend to be aloof from subordinates	I am probably too familiar with subordinates	115
116	I have never tried to teach anybody anything	I think I could do a good job of teaching	116
117	Most of my friends are luckier than I am	I have my share of lucky breaks	117
118	I criticize my subordinates too much	I praise my subordinates too much	118

	L			R	
119	I am stern			I am easy going	119
120	Our neighbors at home were too nosey			Our neighbors went their own way pretty much	120
121	I've had my share of happiness			I've had my share of sickness	121
122	I wish I could have more excitement			I wish I weren't bothered by bad dreams	122
123	The hours at night seem long	Yes	No	....	123
124	I've always been somewhat uneasy			Nothing ever bothers me	124
125	I sometimes drink because it helps me to forget	No	Yes	....	125
126	When I'm excited or upset I don't show it much			When I'm excited or upset I feel sweaty	126
127	Being in a small closed-in place never bothers me			Being in a small closed-in place makes me feel uneasy	127
128	I like most any kind of food			I have a poor appetite	128
129	I've hardly been sick a day in my life			I've had the average amount of sickness	129
130	I have felt bad more from dizziness			I have felt bad more from head cold	130
131	My heart sometimes speeds up for no reason at all	No	Yes	....	131
132	You have to worry more about heart trouble than most people think	No	Yes	....	132
133	I've had my share of backaches, headaches, and stomach aches	Yes	No	....	133
134	I've had a venereal disease	Yes	No	....	134
135	I sometimes faint for no good reason	Yes	No	....	135
136	I have more headaches than the average person	No	Yes	....	136
137	I have occasionally gone to a doctor for headaches or dizzy spells			I have never gone to a doctor for headaches or dizzy spells	137
138	Someone in my family has gone to a doctor or hospital for nervousness	Yes	No	....	138
139	I'm very careful to take medicine whenever I need it			Most people aren't as sick as they think they are	139
140	I'd rather be healthy			I'd rather be well liked	140
141	I'm happier in the morning			I'm happier in the afternoon	141

[redacted]

	L	R	
142	Noise and confusion bother me more	Continued silence bothers me more	142
143	I feel uncomfortable about being in high places	Being up in high places never bothers me	143
144	I'm more afraid of getting shell shocked	I'm more afraid of being physically maimed	144
145	I am more afraid of drowning	I am more afraid of being shot	145
146	I like to have things repeated	I always hear correctly the first time	146
147	I prefer to receive oral orders from my superior	I prefer to receive written orders	147
148	I feel that my education and talents are being wasted right now	I feel that the war is making full use of my talents and education	148
149	The Navy is more deserving of praise than the Merchant Marine Yes	....No	149
150	Most of the things I learn in the service are only going to help win the war	Most of the things I learn in the service will be of value to me when I get out	150
151	I prefer to have good men under me	I prefer to have good officers over me	151
152	Leadership is inborn	Leadership can be learned	152
153	The chief function of an officer is to keep his men working	The chief function of an officer is to guide his men	153
154	It is more important for an officer to know the good points of his men	It is more important for an officer to know the weak points of his men	154
155	It is more important for an officer to feel at ease with other officers	It is more important for an officer to feel at ease with his men	155
156	Officers should welcome advice only from other officers	Officers should welcome advice from the men in their company	156
157	An officer should never be sarcastic	Some men will respond only to sarcasm	157
158	A good leader pitches in	A good leader arranges things so that it isn't necessary for him to pitch in	158
159	Officers should be automatically promoted at scheduled intervals	Promotion should come only as a reward for merit	159
160	In some ways a scapegoat is an asset in a group	Blame can always be correctly placed where it belongs	160
161	A sense of humor relaxes discipline	A sense of humor can aid discipline	161
162	Criticism helps a person do better	Criticism makes a person resentful	162
163	Good morale depends more on good men	Good morale depends more on good equipment	163
164	Some outfits are more important than others in winning the war	Each outfit is equally important in winning the war	164

UNITED STATES NAVY

GENERAL CLASSIFICATION TEST

FORM I

NAVPERS 16500

1943

General Directions

Do not write or make marks of any kind in this test booklet.

Your answers to the questions in this test are to be marked on the separate answer sheet. Use only the special pencil which has been given you for marking your answers. You are to mark your answer to each question by blacking one of the spaces opposite the number of the question. In blacking in the space, make several strokes up and down until you have made a glossy black mark. You may erase a mark if you wish to change an answer, but be sure to erase thoroughly. Remember, make no marks on this test booklet.

Before you start each part of the test, you will have a chance to read special directions for that part and to work some practice problems.

Give only one answer to each question; double answers are counted as incorrect.

Do not spend too much time on any one question. If you are uncertain, make the best choice you can. Do not skip any questions. Work as rapidly and as accurately as you can.

DO NOT OPEN THIS BOOKLET UNTIL YOU ARE TOLD TO DO SO.

[redacted]

PART I

SENTENCE COMPLETION

Directions and Samples

In this test you will be given sentences to read. One word has been left out of each. Following each sentence are printed five words, one of which is the word which was left out. You are to decide which one of the five words belongs in the sentence.

Look at sample sentence A below:

A. The color of milk is ....
1 green 2 red 3 white 4 blue 5 orange

The word which was left out of the sentence is white, which is number 3. Therefore the space under 3 has been blacked in on the separate answer sheet opposite A, in the section labeled "COMPLETION -- Samples".

Here is another sample sentence. Find the word which fits in the sentence and black in the proper space on the answer sheet opposite B. Do this now.

B. We looked at the clock to see what .... it was.
1 day 2 year 3 weather 4 town 5 time

The word which fits in best is number 5, time; so you should have blacked in the space under 5 in line B on the answer sheet.

Here are three more sample sentences. Answer them, and mark your answers on the answer sheet.

C. I thought he was asleep because his eyes were ....
1 dark 2 shut 3 dull 4 gray 5 heavy

D. A good sailor will .... the orders of his superior officers.
1 see 2 fear 3 read 4 obey 5 like

E. He awoke early in the morning and saw the .... sinking below the horizon in the west.
1 sun 2 moon 3 storm 4 waves 5 clouds

Now check your answers as I give the correct ones to you.

ARE THERE ANY QUESTIONS?

If you made any mistakes, correct them by thoroughly erasing the wrong answers and marking the answer sheet correctly.

On the following pages you will find more sentences like the ones you have just answered; but some will be more difficult. You should always choose the one word which best fits in with all parts of the sentence.

Mark only one answer to each question. If you are not sure of the answer, make the best choice you can. Do not skip any questions. Work as fast and as accurately as you can.

STOP HERE. DO NOT GO TO THE NEXT PAGE
UNTIL THE SIGNAL "READY BEGIN" IS GIVEN.

[redacted]

SENTENCE COMPLETION

Test Questions

Mark <u>only</u> on separate answer sheet.

1. A boat floats on the ....
   1 water 2 sail 3 oar 4 wharf 5 wind

2. A seaman should .... his duties.
   1 decide 2 explain 3 find 4 know 5 divide

3. Always .... the salute of those under you.
   1 approve 2 seek 3 appreciate 4 watch 5 return

4. Most landlubbers are .... during their first ocean voyage.
   1 seasick 2 frightened 3 dissatisfied 4 drowsy 5 amazed

5. After a six months' voyage, the ship sailed safely into the ....
   1 wharf 2 sea 3 port 4 foam 5 sunset

6. The ship is still missing though we have tried for months to .... it.
   1 see 2 restore 3 search 4 locate 5 save

7. When the wind sprang up, the men in the small boat hoisted the ....
   1 flag 2 sail 3 oars 4 mast 5 signal

8. The men displayed great .... in the face of heavy gun fire.
   1 ability 2 satisfaction 3 agility 4 self-control 5 alacrity

9. He still retains his enthusiasm and courage despite the .... experiences he has had.
   1 varied 2 fatal 3 numerous 4 adventurous 5 harrowing

10. Never stow away a coil of rope unless it is perfectly ...., since it will deteriorate quickly if allowed to remain damp.
    1 slack 2 waxed 3 dry 4 twined 5 intact

11. He found that it did not pay to be .... in disciplinary cases, since the men respected only a commanding officer who was strict with them.
    1 harsh 2 lenient 3 unjust 4 wrong 5 fair

12. The fog that sprang up was so dense that although it was .... he could hardly see his hand before his face.
    1 twilight 2 dawn 3 noon 4 night 5 overcast

13. The torpedo is a short-range weapon; to use it the destroyer must get relatively close to its ....
    1 target 2 base 3 fleet 4 explosion 5 tender

14. It was clear in 1942 that victory over Japan would be an .... victory indeed if it were coupled with a United Nations defeat in Europe at the hands of Germany.
    1 important 2 appalling 3 empty 4 officious 5 indirect

15. Though particular branches of science have .... appeared to be completely investigated and definitely formulated, this has always proved to be an illusion.
    1 once 2 occasionally 3 never 4 always 5 meanwhile

16. The motor torpedo boats are not even big enough to have ....; the one I went out in was known merely as No. 18.
    1 docks 2 turbines 3 reinforcements 4 lifeboats 5 names

DO NOT STOP. GO ON TO THE NEXT PAGE.

17. Any material placed in the boat must be allowed for by .... the number of passengers.
1 ascertaining 2 increasing 3 reducing 4 accommodating 5 predicting

18. We patrolled the Channel, skirting the edges of the mine fields and all the time listening for submarines on the .... device.
1 signalling 2 control 3 aerial 4 detecting 5 depth

19. A vessel navigating in a fog must go no faster than will permit her to .... within the distance she can see ahead.
1 start 2 stop 3 reverse 4 sail 5 hear

20. The more closely the city and the country are knit together the better; their interests in the last analysis are ....
1 practical 2 distinct 3 unselfish 4 necessary 5 mutual

21. It was noticeable that elderly men predominated among the crews; some of the recent recruits, however, were ....
1 inexperienced 2 relaxed 3 active 4 boys 5 enthusiasts

22. A strategist who adheres inflexibly to any set of preconceived rules is hardly likely to be victor against a .... opponent.
1 resourceful 2 persevering 3 vigilant 4 practiced 5 niggardly

23. Even if one insists on regarding the air forces as a separate branch of military power it is a rather .... outlook which concentrates upon the spectacle of the airplane over its target, to the exclusion of the long chain of circumstances which are responsible for putting it there.
1 indefinite 2 normal 3 realistic 4 limited 5 cynical

24. The chief disadvantage of the hydrophone was that it picked up other .... than that of the submarine's propellers.
1 elements 2 features 3 noises 4 clues 5 details

25. In a pursuit as .... as is the waging of war, it might be set down as an axiom that no one factor, and certainly no one weapon, can be exclusively decisive.
1 difficult 2 complex 3 confused 4 challenging 5 painstaking

26. In recent battleships, the proportion of .... devoted to armor has gone above the 40 per cent mark.
1 paint 2 iron 3 tonnage 4 construction 5 space

27. Those to whom the word "defensive" has now become malodorous must be reminded that simultaneous offensives against two powerful and widely removed enemies are usually impossible and always ....
1 accidental 2 hazardous 3 overwhelming 4 interdependent 5 suicidal

28. Despite the great emphasis on protection in the battleship it is armament and not armor that makes a .... ship.
1 light 2 fast 3 heavy 4 manageable 5 fighting

29. It is worth noting that armor decks were used on warships long before the development of the airplane, and even now the heaviest impact which the armor decks are called upon to bear is from ....
1 bombs 2 torpedoes 3 shells 4 depth charges 5 submarines

30. It is difficult to conceive of gunfire at a target thirteen miles away being more .... than aerial bombing from only 2,000 feet elevation, but this is actually the case.
1 accurate 2 inaccurate 3 harmless 4 risky 5 visible

STOP. WAIT FOR FURTHER INSTRUCTIONS.

PART II

O P P O S I T E S

Directions and Samples

Look at this sample question:

F. HOT
1 flat 2 cold 3 light 4 soft 5 wet

You know that the opposite of HOT is cold. Cold is number 2, so number 2 has been blacked in on the answer sheet. Look at your answer sheet, in the section labeled "OPPOSITES -- Samples", to see how this has been done. Notice that the space under 2 has been blacked in opposite F.

In each question in this test you will see a word printed in CAPITAL LETTERS. This word is followed by five numbered words, only one of which means the opposite of the word in capital letters. In each case, you are to find the one word which means the opposite of the word in capital letters.

Here is another practice question. Find the answer, and black in the proper space in line G on the answer sheet.

G. BUSY
1 tired 2 friendly 3 stupid 4 idle 5 weak

The word which means the opposite of BUSY is idle. Idle is number 4, so in line G on the answer sheet, you should have blacked in the space under number 4.

Here are three more practice questions. Do them, and mark your answers on the answer sheet.

H. SELFISH
1 generous 2 lazy 3 playful 4 happy 5 beautiful

I. BREAK
1 weep 2 find 3 handle 4 use 5 mend

J. GRIEF
1 anger 2 poverty 3 joy 4 sorrow 5 pride

Now check your answers while I give the correct ones to you.

ARE THERE ANY QUESTIONS?

If you made any mistakes, erase your wrong answers completely, and then mark your sheet correctly.

In the test on the following pages you will have more questions like the ones you have just done. Answer each question carefully, but if you are not sure, make the best guess that you can.

If you should finish before time is called, go back and check your answers.

DO NOT TURN THIS PAGE UNTIL THE SIGNAL "READY BEGIN" IS GIVEN.

[redacted]

OPPOSITES

Test Questions

Mark only on separate answer sheet.

31. QUICK
    1 strong   2 slow   3 dry   4 dark   5 neat

32. STRAIGHT
    1 shiny   2 low   3 crooked   4 clean   5 long

33. CHEERFUL
    1 clumsy   2 wise   3 active   4 plain   5 gloomy

34. GIVE
    1 lose   2 keep   3 hear   4 live   5 hate

35. SHARP
    1 hollow   2 blunt   3 calm   4 heavy   5 clear

36. TRUST
    1 blame   2 deny   3 forget   4 doubt   5 surprise

37. REAL
    1 foreign   2 imaginary   3 empty   4 faint   5 rare

38. JOIN
    1 mislay   2 confuse   3 refute   4 distrust   5 separate

39. CONCEITED
    1 afraid   2 rough   3 modest   4 quiet   5 ashamed

40. ADEQUATE
    1 improper   2 deceitful   3 insufficient   4 unprotected   5 unnecessary

41. DISCLOSE
    1 conceal   2 befriend   3 rectify   4 observe   5 suspect

42. COMPETENT
    1 dissatisfied   2 corrupt   3 unfriendly   4 incapable   5 unopposed

43. COMPLIANCE
    1 exposure   2 resistance   3 inference   4 reprisal   5 perseverance

44. POLLUTE
    1 determine   2 sustain   3 convert   4 dilute   5 purify

45. CONCENTRATE
    1 disapprove   2 cheapen   3 diffuse   4 conflict   5 refuse

46. RIGOROUS
    1 delinquent   2 superstitious   3 voluntary   4 refined   5 lenient

DO NOT STOP. GO ON TO THE NEXT PAGE.

47. OCCASIONAL
1 frequent 2 former 3 manifold 4 concurrent 5 eventual

48. ACCELERATE
1 punish 2 grovel 3 release 4 soothe 5 retard

49. INVOLVE
1 allay 2 derange 3 suppress 4 exclude 5 dispel

50. IMPAIR
1 enhance 2 damage 3 complement 4 sanction 5 acquiesce

51. INDISPENSABLE
1 compensatory 2 superfluous 3 exceptional 4 reprehensible
5 conditional

52. METICULOUS
1 slovenly 2 prudish 3 indiscreet 4 confiding 5 gluttonous

53. CONCERTED
1 unmusical 2 discouraged 3 inapplicable 4 disorganized 5 mistaken

54. VINDICTIVE
1 irascible 2 unselfish 3 forgiving 4 reactive 5 communicative

55. LATENT
1 stunted 2 beneficent 3 secular 4 apparent 5 mismanaged

56. BLATANT
1 astute 2 deliberate 3 jubilant 4 reticent 5 melancholy

57. DISPARATE
1 reciprocal 2 moderate 3 appropriate 4 concomitant 5 identical

58. EQUABLE
1 unjustified 2 bewildering 3 unstable 4 eccentric 5 callous

59. INIMITABLE
1 fabulous 2 ordinary 3 jocund 4 loquacious 5 mitigated

60. RECANT
1 avow 2 congeal 3 covet 4 expunge 5 elicit

STOP. WAIT FOR FURTHER INSTRUCTIONS.

PART III

A N A L O G I E S

Directions and Samples

In this test you will be asked to find certain relationships between words. You will have a number of sentences to read; in each sentence you will have one pair of words, such as NIGHT and dark, which have the same relationship as another pair of words, such as DAY and light.

Read the following sentence:
During the NIGHT it is dark, but during the DAY it is light.
Notice that "night" is related to "dark" in the same way that "day" is related to "light", so we may say that

NIGHT is to dark as DAY is to light.

Now look at this one:
You READ a book and LISTEN to music.
Here you can see that

READ is to book as LISTEN is to music.

In the question labeled K below, you are to select the one of the five numbered words which best completes the thought.

K. READ is to book as LISTEN is to
1 music   2 air   3 pencil   4 newspaper   5 platform

The correct answer is "music", or number 1. Notice on the answer sheet, in the section labeled "ANALOGIES -- Samples", that space 1 has been blacked in, opposite the letter K.

What is the correct answer to the following question?

L. GLOVE is to hand as HAT is to
1 face   2 fingers   3 forehead   4 body   5 head

A GLOVE is worn on the hand, and a HAT is worn on the head; so head, or number 5, is the correct answer. Show the answer by blacking in the space under 5, opposite the letter L, on the answer sheet. Do this now.

Find the answers to the following questions yourself; record your answers on the answer sheet.

M. FISH is to swim as BIRD is to
1 air   2 fly   3 feathers   4 ride   5 wings

N. WATER is to sponge as INK is to
1 pen   2 bottle   3 write   4 blotter   5 desk

O. THERMOMETER is to temperature as CLOCK is to
1 month   2 heat   3 time   4 hand   5 dial

DO NOT STOP. GO ON TO THE NEXT PAGE.

Now check your answers as I give the correct ones to you.

ARE THERE ANY QUESTIONS?

If you made any mistakes, erase your wrong answers completely, and then mark your sheet correctly.

In the test on the following pages you will have more questions like the ones you have just done. Answer each question carefully, but if you are not sure, make the best guess that you can.

If you should finish before time is called, go back and check your answers.

DO NOT TURN THIS PAGE UNTIL THE SIGNAL
"READY BEGIN" IS GIVEN.

[redacted]

ANALOGIES

Mark only on separate answer sheet.

61. EYE is to see as EAR is to
 1 nose 2 head 3 hear 4 speak 5 listen

62. TREE is to branch as HAND is to
 1 arm 2 glove 3 wrist 4 finger 5 fist

63. DOOR is to room as LIPS are to
 1 tongue 2 mouth 3 nose 4 teeth 5 face

64. GUN is to aim as BOAT is to
 1 rudder 2 float 3 steer 4 sail 5 navigate

65. FISH is to scales as BIRD is to
 1 wings 2 feet 3 eggs 4 feathers 5 skin

66. HOUSE is to foundation as TREE is to
 1 garden 2 branches 3 roots 4 bark 5 leaves

67. SHIP is to lifeboat as PLANE is to
 1 safety belt 2 hangar 3 aircraft carrier 4 glider 5 parachute

68. FOREST is to trees as NAVY is to
 1 sailors 2 admirals 3 defense 4 food 5 ocean

69. CROP is to land as FISH is to
 1 fisherman 2 sea 3 market 4 fish-net 5 boat

70. AIR is to wind as WATER is to
 1 land 2 waves 3 clouds 4 ocean 5 fish

71. ASK is to receive as ATTEMPT is to
 1 desire 2 give 3 begin 4 wait 5 succeed

72. TRAINS are to passengers as BOOKS are to
 1 book-stores 2 authors 3 paper 4 publishers 5 ideas

73. THROW is to catch as TEACH is to
 1 remember 2 study 3 learn 4 read 5 respect

74. GROUND is to worm as OCEAN is to
 1 submarine 2 destroyer 3 cable 4 sea 5 depth bomb

75. LAVA is to volcano as SPEECH is to
 1 words 2 orator 3 audience 4 eloquence 5 temper

76. SALARY is to service as PENALTY is to
 1 work 2 disapproval 3 discipline 4 misdemeanor 5 pain

77. MATCH is to flame as DETONATOR is to
 1 explosion 2 trigger 3 bullet 4 gun 5 powder

78. WATER is to island as LAND is to
 1 continent 2 peninsula 3 ocean 4 lake 5 stream

79. TROUBLE is to unhappiness as SUN is to
 1 heavens 2 earth 3 warmth 4 cold 5 moon

 DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]

80. LITERATURE is to book as ART is to
1 artist 2 painting 3 color 4 beauty 5 scenery

81. INVENTION is to idea as RIVER is to
1 ocean 2 flood 3 crops 4 spring 5 bank

82. POSITIVE is to uncertain as KNOWLEDGE is to
1 logic 2 faith 3 proof 4 superstition 5 convention

83. CREST is to trough as HILL is to
1 mountain 2 meadow 3 slope 4 lake 5 valley

84. AIRPLANE is to wings as AUTOMOBILE is to
1 motor 2 land 3 wheels 4 brake 5 hood

85. INTERMITTENT is to continuous as FLICKER is to
1 frighten 2 blink 3 fade 4 glare 5 glint

86. THREATEN is to fulfill as CLOUDS are to
1 winter 2 darkness 3 storm 4 dampness 5 sunshine

87. TWILIGHT is to darkness as SORROW is to
1 misfortune 2 despair 3 discomfort 4 happiness 5 regret

88. MORNING is to wake as SPRING is to
1 flowers 2 thaw 3 grow 4 sprout 5 welcome

89. MONTH is to time as QUART is to
1 gallon 2 measure 3 volume 4 peck 5 scales

90. CHARACTER is to reputation as REALITY is to
1 appearance 2 conduct 3 integrity 4 wisdom 5 possibility

91. STREAM is to torrent as RETREAT is to
1 rout 2 victory 3 failure 4 army 5 battle

92. TELESCOPE is to lens as VIOLIN is to
1 violinist 2 bow 3 strings 4 music 5 piano

93. LIGHTHOUSE is to reef as CONSCIENCE is to
1 man 2 conduct 3 good 4 soul 5 temptation

94. CONTINENT is to isthmus as OCEAN is to
1 channel 2 bay 3 peninsula 4 strait 5 lake

95. BLOTTER is to ink as EGOTIST is to
1 friends 2 enemies 3 selfishness 4 contempt 5 praise

96. DIVE is to fall as SWIM is to
1 sink 2 float 3 drown 4 rise 5 rescue

97. MEDICINES are to health as FERTILIZERS are to
1 riches 2 soil 3 agriculture 4 farm 5 productivity

98. MATCH is to fuse as FUSE is to
1 explosion 2 gun 3 bullet 4 powder 5 trigger

99. PEN is to ink as SUBMARINE is to
1 torpedo 2 explosion 3 dive bomber 4 ocean 5 ship

100. SAPLING is to lumber as CALF is to
1 cow 2 beef 3 shoes 4 bull 5 veal

STOP. WAIT FOR FURTHER INSTRUCTIONS.

[redacted]

UNITED STATES NAVY

G E N E R A L   C L A S S I F I C A T I O N   T E S T

FORM 2

NAVPERS - 16502

1943

General Directions

Do not write or make marks of any kind in this test booklet.

Your answers to the questions in this test are to be marked on the separate answer sheet. Use only the special pencil which has been given you for marking your answers. You are to mark your answer to each question by blacking one of the spaces opposite the number of the question. In blacking in the space, make several strokes up and down until you have made a glossy black mark. You may erase a mark if you wish to change an answer, but be sure to erase thoroughly. Remember, make no marks on this test booklet.

Before you start each part of the test, you will have a chance to read special directions for that part and to work some practice problems.

Give only one answer to each question; double answers are counted as incorrect.

Do not spend too much time on any one question. If you are uncertain, make the best choice you can. Do not skip any questions. Work as rapidly and as accurately as you can.

NOTICE:
This test is not to be shown, or the contents revealed, to unauthorized persons in or out of the Navy, or reproduced in whole or in part without written authorization from the Bureau of Naval Personnel of the United States Navy. This test booklet must be surrendered with the answer sheet at the close of the test.

DO NOT OPEN THIS BOOKLET UNTIL YOU ARE TOLD TO DO SO.

[redacted]

PART I

S E N T E N C E   C O M P L E T I O N

Directions and Samples

In this test you will be given sentences to read. One word has been left out of each. Following each sentence are printed five words, one of which is the word which was left out. You are to decide which one of the five words belongs in the sentence.

Look at sample sentence A below:

A. The color of milk is ....
   1 green   2 red   3 white   4 blue   5 orange

The word which was left out of the sentence is white, which is number 3. Therefore the space under 3 has been blacked in on the separate answer sheet opposite A, in the section labeled "COMPLETION -- Samples".

Here is another sample sentence. Find the word which fits in the sentence and black in the proper space on the answer sheet opposite B. Do this now.

B. We looked at the clock to see what .... it was.
   1 day   2 year   3 weather   4 town   5 time

The word which fits in best is number 5, time; so you should have blacked in the space under 5 in line B on the answer sheet.

Here are three more sample sentences. Answer them, and mark your answers on the answer sheet.

C. I thought he was asleep because his eyes were ....
   1 dark   2 shut   3 dull   4 gray   5 heavy

D. A good sailor will .... the orders of his superior officers.
   1 see   2 fear   3 read   4 obey   5 like

E. He awoke early in the morning and saw the .... sinking below the horizon in the west.
   1 sun   2 moon   3 storm   4 waves   5 clouds

Now check your answers as the correct ones are given to you.

ARE THERE ANY QUESTIONS?

If you made any mistakes, correct them by thoroughly erasing the wrong answers and marking the answer sheet correctly.

On the following pages you will find more sentences like the ones you have just answered; but some will be more difficult. You should always choose the one word which best fits in with all parts of the sentence.

Mark only one answer to each question. If you are not sure of the answer, make the best choice you can. Do not skip any questions. Work as rapidly and as accurately as you can.

 DO NOT GO ON TO THE NEXT PAGE UNTIL THE SIGNAL "READY - BEGIN" IS GIVEN.

 [redacted]

SENTENCE COMPLETION

Test Questions

Mark <u>only</u> on separate answer sheet.

1. The main task of a navy is to win ....
   1 ships 2 guns 3 prizes 4 friends 5 battles

2. It takes a long time to .... a battleship.
   1 build 2 defend 3 float 4 stop 5 attack

3. Strong electric currents can have such a .... effect on the body that the injury they cause may be permanent.
   1 mysterious 2 harmful 3 healthful 4 thrilling 5 mild

4. A navy is large if it has many ....
   1 duties 2 officers 3 oceans 4 ships 5 parts

5. When a steam and a sailing vessel are proceeding so as to involve risk of .... , the steam vessel shall keep out of the way of the sailing vessel.
   1 upset 2 collision 3 sinking 4 grounding 5 stopping

6. The air escaping from the old type of torpedo as it traveled through the water marked its .... so that it could sometimes be avoided by its intended victim.
   1 depth 2 path 3 size 4 source 5 distance

7. The men must learn to .... fresh water as much as possible, as it is obtained at great expense on board ship.
   1 purify 2 drink 3 conserve 4 enjoy 5 secure

8. Because the men had taken their shoes off, their feet became .... from wading up onto the stony beach.
   1 tired 2 wet 3 hot 4 cold 5 bruised

9. A few torpedo hits on important ships of a retreating fleet may force its admiral into an action he would otherwise ....
   1 seek 2 win 3 avoid 4 lose 5 dislike

10. The .... of a bullet shot from a modern rifle is such that the bullet may strike a man before he hears the sound of the explosion which sent it on its way.
    1 impact 2 path 3 speed 4 effect 5 sound

11. The captain doubted whether the submarine's ballast tanks were still intact; if they were not, there was no hope of blowing the water out of them and thereby losing enough .... to allow the boat to return to the surface.
    1 weight 2 air 3 speed 4 bulk 5 descent

12. Supplies were getting scarce and the torpedo boats were .... to make further offensive raids since we couldn't afford to risk the boats or spare the gas.
    1 forbidden 2 needed 3 forced 4 prepared 5 used

13. In naval warfare, the side which sees that it is clearly weaker will usually attempt to .... action.
    1 begin 2 evade 3 win 4 settle 5 organize

14. Sailing vessels may seem to be very .... to operate, but due to the cost of handling sails they are more expensive than steam-driven freighters.
    1 difficult 2 easy 3 cheap 4 slow 5 costly

15. TNT is a powerful and dangerous explosive, but it can be handled .... provided its characteristics are known.
    1 carefully 2 skillfully 3 cheaply 4 speedily 5 safely

DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]

16. The correct way to remove the insulation from heavy wire is to .... the insulation off the wire as you would sharpen a lead pencil.
1 peel 2 whittle 3 pull 4 pick 5 rub

17. Although the ship was very .... , he could see every detail because the air was unusually clear.
1 close 2 vague 3 large 4 distant 5 indistinct

18. The importance in warfare of ability to transport supplies is generally recognized; however, upon the entrance of Japan and the United States into the war the daily press was full of comparisons of the naval strength of the two countries but entirely neglected the matter of .... tonnage.
1 lost 2 portable 3 shipping 4 heavy 5 utilizing

19. On board ship, the .... of time is absolutely necessary because it may be impossible to determine a ship's position unless the time of day is known.
1 planning 2 passage 3 saving 4 measurement 5 minimum

20. The .... of the explosive charge in a shell is not a good indication of its destructive power because some explosives are much more powerful pound for pound than others.
1 structure 2 weight 3 size 4 danger 5 appearance

21. Much of the criticism of the government arose because surface developments were taken as indications of basic policy when they were in fact merely .... steps in the process of achieving that policy.
1 harmless 2 preliminary 3 general 4 alleged 5 controversial

22. There is a saturation point of supply, and all above that .... successful operations.
1 determines 2 prolongs 3 hinders 4 paralyzes 5 eases

23. Despite the defenders' resort to the "scorched earth" policy in Malaya, the Japanese .... vast stores of tin and rubber with possibilities of later exploitation.
1 attacked 2 had 3 destroyed 4 maintained 5 acquired

24. Wire rope of the same size as Manila rope is much stronger, so that for a given task a wire rope may be used which is .... than the Manila rope which would be needed for the same purpose.
1 thinner 2 tougher 3 thicker 4 stiffer 5 stronger

25. The formation of rust is .... by heat, as is shown by the fact that the parts of a ship's hull near the boilers rust more rapidly than other parts.
1 damaged 2 accompanied 3 delayed 4 increased 5 caused

26. In the majority of cases gas poisoning is due to carbon monoxide, and unless it is known definitely that some other poisonous gas is responsible, it is usually .... that carbon monoxide is the cause of asphyxiation.
1 known 2 clear 3 found 4 supposed 5 doubted

27. Care is sometimes necessary when surprise of the enemy is needed, but at other times caution may actually result in greater .... than would boldness.
1 losses 2 victories 3 advantages 4 effectiveness 5 bravery

28. The great commander must be able to stick to his course despite a thousand distractions, and yet be sufficiently .... to recognize when a change in circumstances demands a change in plan.
1 agreeable 2 informed 3 flexible 4 independent 5 bold

29. Intensified enemy submarine operations in the western Atlantic soon put increased strain upon shipping that had already been .... by a greater military demand for transports.
1 taxed 2 expanded 3 weakened 4 necessitated 5 required

30. The propaganda expert is convinced that the psychological front is decisive in modern warfare, the economist is equally certain that the production effort is most important, and among military men there will be all kinds of divisions of opinion according to the .... of the disputants.
1 wisdom 2 characteristics 3 ambitions 4 intelligence 5 specialties

STOP. WAIT FOR FURTHER INSTRUCTIONS.

[redacted]

PART II

O P P O S I T E S

Directions and Samples

Look at this sample question:

F. HOT
   1 flat   2 cold   3 light   4 soft   5 wet

You know that the opposite of HOT is cold. Cold is number 2, so number 2 has been blacked in on the answer sheet. Look at your answer sheet, in the section labeled "OPPOSITES -- Samples," to see how this has been done. Notice that the space under 2 has been blacked in opposite F.

In each question in this test you will see a word printed in CAPITAL LETTERS. This word is followed by five numbered words, only one of which means the opposite of the word in capital letters. In each case, you are to find the one word which means the opposite of the word in capital letters.

Here is another practice question. Find the answer, and black in the proper space in line G on the answer sheet.

G. BUSY
   1 tired   2 friendly   3 stupid   4 idle   5 weak

The word which means the opposite of BUSY is idle. Idle is number 4, so in line G on the answer sheet, you should have blacked in the space under number 4.

Here are three more practice questions. Do them, and mark your answers on the answer sheet.

H. SELFISH
   1 generous   2 lazy   3 playful   4 happy   5 beautiful

I. BREAK
   1 weep   2 find   3 handle   4 use   5 mend

J. GRIEF
   1 anger   2 poverty   3 joy   4 sorrow   5 pride

Now check your answers as the correct ones are given to you.

ARE THERE ANY QUESTIONS?

If you made any mistakes, erase your wrong answers completely, and then mark your sheet correctly.

In the test on the following pages you will have more questions like the ones you have just done. Answer each question carefully, but if you are not sure, make the best guess that you can.

If you should finish before time is called, go back and check your answers.

DO NOT TURN THIS PAGE UNTIL THE SIGNAL "READY - BEGIN" IS GIVEN.

[redacted]

OPPOSITES

Test Questions

Mark <u>only</u> on separate answer sheet.

31. EARLY
    1 past   2 slow   3 lazy   4 late   5 future

32. ANGRY
    1 pleased   2 hopeful   3 careful   4 comfortable   5 boastful

33. FIND
    1 send   2 leave   3 lose   4 drop   5 hide

34. APPROACH
    1 extend   2 dismiss   3 attack   4 withdraw   5 rebel

35. FASTENED
    1 abandoned   2 loose   3 wide   4 broken   5 exposed

36. INTERESTING
    1 serious   2 unpopular   3 dull   4 difficult   5 foolish

37. MATURE
    1 undeveloped   2 partial   3 weak   4 inferior   5 ridiculous

38. PROBABLE
    1 possible   2 unsuitable   3 concealed   4 unlikely   5 unfavorable

39. COMPULSORY
    1 superfluous   2 voluntary   3 enthusiastic   4 indulgent   5 desired

40. GRADUAL
    1 unexpected   2 irregular   3 greedy   4 excessive   5 sudden

41. YIELD
    1 intend   2 repudiate   3 provoke   4 manage   5 withstand

42. EXCLUDE
    1 connect   2 pamper   3 admit   4 repay   5 support

43. BRISK
    1 sluggish   2 coarse   3 vague   4 shy   5 smooth

44. FACILITATE
    1 hinder   2 obey   3 camouflage   4 coerce   5 clarify

DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]

45. SYNTHETIC
1 moist 2 intense 3 crude 4 normal 5 natural

46. INTRICATE
1 neat 2 simple 3 complex 4 vast 5 soiled

47. DEPLETE
1 acquire 2 pump 3 diffuse 4 replenish 5 value

48. PROMINENT
1 unessential 2 pitiful 3 inconspicuous 4 contemptible 5 futile

49. RIGID
1 curved 2 crooked 3 pliant 4 chaotic 5 floating

50. ANONYMOUS
1 identified 2 authentic 3 egotistic 4 meaningful 5 outstanding

51. STRINGENT
1 lax 2 muffled 3 subtle 4 embellished 5 fragrant

52. PRECARIOUS
1 secure 2 bold 3 convincing 4 deep 5 weighty

53. PRECEDE
1 proceed 2 depart 3 oppose 4 follow 5 claim

54. SUBSEQUENT
1 ultimate 2 prior 3 retarded 4 superb 5 derivative

55. CESSATION
1 encroachment 2 variation 3 continuation 4 creation 5 acquisition

56. CIRCUITOUS
1 concrete 2 conclusive 3 brief 4 direct 5 compact

57. UNPRECEDENTED
1 awaited 2 customary 3 premature 4 resultant 5 basic

58. INCENTIVE
1 contradiction 2 proposal 3 requirement 4 expenditure 5 deterrent

59. LUCRATIVE
1 unprofitable 2 parsimonious 3 emollient 4 tedious 5 economical

60. AUGMENT
1 conciliate 2 pollute 3 diminish 4 repress 5 exhaust

STOP. WAIT FOR FURTHER INSTRUCTIONS.

PART III

A N A L O G I E S

Directions and Samples

In this test you will be asked to find certain relationships between words. You will have a number of sentences to read; in each sentence you will have one pair of words, such as NIGHT and dark, which have the same relationship as another pair of words, such as DAY and light.

Read the following sentence:

During the NIGHT it is dark, but during the DAY it is light.

Notice that "night" is related to "dark" in the same way that "day" is related to "light", so we may say that

NIGHT is to dark as DAY is to light.

Now look at this one:

You READ a book and LISTEN to music.

Here you can see that

READ is to book as LISTEN is to music.

In the question labeled K below, you are to select the one of the five numbered words which best completes the thought.

K. READ is to book as LISTEN is to
1 music 2 air 3 pencil 4 newspaper 5 platform

The correct answer is "music", or number 1. Notice on the answer sheet, in the section labeled "ANALOGIES -- Samples", that space 1 has been blacked in, opposite the letter K.

What is the correct answer to the following question?

L. GLOVE is to hand as HAT is to
1 face 2 fingers 3 forehead 4 body 5 head

A GLOVE is worn on the hand, and a HAT is worn on the head; so head, or number 5, is the correct answer. Show the answer by blacking in the space under 5, opposite the letter L, on the answer sheet. Do this now.

Find the answers to the following questions yourself; record your answers on the answer sheet.

M. FISH is to swim as BIRD is to
1 air 2 fly 3 feathers 4 ride 5 wings

N. WATER is to sponge as INK is to
1 pen 2 bottle 3 write 4 blotter 5 desk

O. THERMOMETER is to temperature as CLOCK is to
1 month 2 heat 3 time 4 hand 5 dial

Now check your answers as I give the correct ones to you.

ARE THERE ANY QUESTIONS?

If you made any mistakes, erase your wrong answers completely, and then mark your sheet correctly.

In the test on the following pages you will have more questions like the ones you have just done. Answer each question carefully, but if you are not sure, make the best guess that you can.

If you should finish before time is called, go back and check your answers.

DO NOT TURN THIS PAGE UNTIL THE SIGNAL "READY - BEGIN" IS GIVEN.

ANALOGIES

Test Questions

Mark only on separate answer sheet.

61. AIRPLANE is to air as BOAT is to
    1 fish 2 water 3 sail 4 oar 5 captain

62. AIR is to breathe as WATER is to
    1 wash 2 swim 3 drink 4 flow 5 dive

63. TRACK is to train as ROAD is to
    1 automobile 2 country 3 tractor 4 concrete 5 travel

64. MAN is to food as MOTOR is to
    1 automobile 2 gasoline 3 starter 4 grease 5 driver

65. ERASER is to pencil mark as SOAP is to
    1 bath 2 dirt 3 water 4 cleanliness 5 skin

66. CREW is to men as FLEET is to
    1 destroyers 2 officers 3 seas 4 ships 5 seapower

67. TACK is to nail as NAIL is to
    1 spike 2 hammer 3 bolt 4 wedge 5 screw

68. SEEP is to gush as SPRINKLE is to
    1 drown 2 pour 3 dip 4 tumble 5 overflow

69. GIANT is to dwarf as BOULDER is to
    1 man 2 rock 3 mountain 4 sand 5 pebble

70. SAPLING is to tree as CHILD is to
    1 adult 2 growth 3 maturity 4 youth 5 infant

71. FEAT is to self-satisfaction as BLUNDER is to
    1 silence 2 excuse 3 grumbling 4 embarrassment 5 ridicule

72. BATTLE is to Navy as GAME is to
    1 player 2 college 3 skill 4 competition 5 team

73. BLOCKADE is to commerce as TOURNIQUET is to
    1 limb 2 wound 3 bleeding 4 first aid 5 bandages

74. TRUCK is to utility as LIMOUSINE is to
    1 wealth 2 transportation 3 ornamentation 4 expense 5 luxury

75. HAND is to finger as FOOT is to
    1 shoe 2 ground 3 ankle 4 leg 5 toe

76. UMBRELLA is to rain as HELMET is to
    1 head 2 shrapnel 3 poison gas 4 protection 5 firearms

77. HOUSE is to architect as AUTOMOBILE is to
    1 mechanic 2 manufacturer 3 owner 4 designer 5 driver

78. WILDERNESS is to houses as CONTENTMENT is to
    1 joy 2 poverty 3 troubles 4 pleasures 5 duties

79. CUT is to break as EXACT is to
    1 distorted 2 mistaken 3 disjoined 4 approximate 5 unskillful

DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]

80. WAGES is to work as IMPRISONMENT is to
1 criminal 2 justice 3 crime 4 reform 5 punishment

81. ILLNESS is to symptom as GUILT is to
1 proof 2 evidence 3 misdeed 4 witness 5 conscience

82. EXPERIENCE is to memory as INJURY is to
1 cure 2 salve 3 pain 4 scar 5 revenge

83. POLL TAX is to voting as TARIFF is to
1 government 2 industry 3 taxation 4 amusements 5 imports

84. WATER is to thirst as SUCCESS is to
1 ambition 2 theory 3 failure 4 ability 5 opportunity

85. PERFORMANCE is to applause as CHARACTER is to
1 respect 2 training 3 actions 4 appearance 5 self-esteem

86. HOE is to weeds as CANNON is to
1 bullets 2 battle 3 enemy 4 battlefield 5 destruction

87. TEACHING is to ignorance as FIRE is to
1 cold 2 heat 3 fuel 4 temperature 5 boiler

88. FLUID is to strainer as NEWS is to
1 public 2 opinion 3 analysis 4 censorship 5 newspaper

89. BUILDING is to plans as BOOK is to
1 outline 2 illustrations 3 preface 4 topic 5 review

90. FAMILY is to home as FLEET is to
1 ocean 2 harbor 3 base 4 flagship 5 native land

91. BREAKDOWN is to repairs as DELINQUENCY is to
1 rehabilitation 2 punishment 3 amnesty 4 obedience 5 integrity

92. SOW is to reap as PRACTICE is to
1 exert 2 compete 3 finish 4 prepare 5 excel

93. MATCH is to flame as TALENT is to
1 effort 2 achievement 3 fame 4 training 5 ability

94. SUCCEED is to opportunity as ENTER is to
1 encouragement 2 room 3 key 4 trickery 5 challenge

95. MONTH is to time as INCH is to
1 ruler 2 measure 3 quantity 4 foot 5 length

96. NEATNESS is to disorder as GOVERNMENT is to
1 freedom 2 citizens 3 opposition 4 anarchy 5 fascism

97. FUEL is to fire as ENERGY is to
1 strength 2 inspiration 3 activity 4 health 5 purpose

98. PRAIRIE is to garden as FOREST is to
1 trees 2 jungle 3 lumber 4 field 5 orchard

99. FABRIC is to thread as KNOWLEDGE is to
1 fact 2 intelligence 3 instruction 4 belief 5 judgment

100. CHICK is to omelette as CALF is to
1 beef 2 hay 3 cream 4 cheese 5 shoes

STOP. WAIT FOR FURTHER INSTRUCTIONS.

[redacted]

UNITED STATES NAVY GENERAL CLASSIFICATION TEST

NAVPERS 16501

Co. No.

FORM No. ______

G. C. T. SCORE

NAME ______ (LAST NAME) ______ (FIRST NAME) ______ (MIDDLE NAME)

AGE ______ LAST GRADE COMPLETED ______

COMPLETION

SAMPLES

	1	2	3	4	5
A			▌		
B					
C					
D					
E					

BEGIN TEST HERE

Items 1–30, each with answer spaces 1 2 3 4 5.

OPPOSITES

SAMPLES

	1	2	3	4	5
F		▌			
G					
H					
I					
J					

BEGIN TEST HERE

Items 31–60, each with answer spaces 1 2 3 4 5.

ANALOGIES

SAMPLES

	1	2	3	4	5
K	▌				
L					
M					
N					
O					

BEGIN TEST HERE

Items 61–100, each with answer spaces 1 2 3 4 5.

Licensed under Patent 1,586,628

IBM FORM I.T.S. 1000 A 1043 REV.

[redacted]

UNITED STATES NAVY

TESTS OF READING
AND
ARITHMETICAL REASONING

FORM I

NAVPERS 16510

1943

NOTICE:

This test is not to be shown, or the contents revealed, to unauthorized persons in or out of the Navy, or reproduced in whole or in part without written authorization from the Bureau of Naval Personnel of the United States Navy. This test booklet must be surrendered with the answer sheet at the close of the test.

DO NOT OPEN THIS BOOKLET UNTIL YOU ARE TOLD TO DO SO.

[redacted]

TEST I

R E A D I N G

Directions and Samples

In this test you will be given several paragraphs to read. Following each paragraph are a number of incomplete sentences, each of which is followed by five phrases lettered from (a) to (e). One of the five phrases added to the incomplete sentence makes a statement which according to the paragraph is true. After reading the paragraph, you are to decide which phrase makes each statement true. Then black in the space on the separate answer sheet which corresponds to that phrase.

Read the following sample paragraph; then answer question A.

After a can of paint has been opened and the paint partly used, the can should be covered and kept as air-tight as possible to prevent a paint scum from forming on the surface. If scum forms, the paint should be strained through a fine-mesh wire screen or cheesecloth.

A. To prevent scum from forming in a partly-used can of paint, one should
- (a) keep the paint free from dirt.
- (b) fill the can up with water.
- (c) stir the paint well before storing.
- (d) keep the can tightly covered.
- (e) make sure that the can is more than half full.

The paragraph tells you that the can should be covered and kept as air-tight as possible to prevent scum from forming; so (d) is the phrase which completes the statement correctly. Therefore (d) has been blacked in on the separate answer sheet opposite A in the section labeled "Reading -- Samples".

Now answer question B and black in the space in line B on the answer sheet which corresponds to the phrase which completes the statement correctly, according to the paragraph.

B. The method of removing scum from paint is
- (a) to stir the paint thoroughly.
- (b) to pour the scum off.
- (c) to strain the paint through a wire screen.
- (d) to skim the scum off with a putty knife.
- (e) not discussed in the paragraph.

Answer question C and black in the space on the answer sheet which corresponds to the correct phrase.

C. The method of thinning paint is
(a) to mix it with turpentine.
(b) to mix it with linseed oil.
(c) to mix it with white lead.
(d) to mix it with water.
(e) not discussed in the paragraph.

Now check your answers while I give the correct ones to you.

ARE THERE ANY QUESTIONS?

When the signal is given, you will have more paragraphs to read and questions to answer. Mark each answer in the proper space on the separate answer sheet. Mark only one answer to each question. Do not make any marks in this booklet. Work as fast and as accurately as you can.

Answer every question. If you are not sure of the answer, make the best guess you can.

DO NOT TURN THIS PAGE UNTIL THE SIGNAL
"READY BEGIN" IS GIVEN

READING

Test Questions

Mark only on separate answer sheet.

Cruisers have light armor, carry guns of moderate size, and are able to travel at high speed. Cruisers whose largest guns are greater than 6 inches are known as heavy cruisers, while those whose largest guns are 6 inches or less are known as light cruisers. All cruisers have very large fuel tanks in order to maintain high speed for a long time. A cruiser is divided into numerous water-tight compartments, so that a hole in one part of the ship will flood only a part of the ship.

1. Cruisers whose largest guns are 6-inch guns are
   (a) light cruisers.
   (b) medium cruisers.
   (c) heavy cruisers.
   (d) auxiliaries.
   (e) not described in the paragraph.

2. Cruisers have
   (a) heavy armor, high speed, and large guns.
   (b) light armor, low speed, and medium guns.
   (c) light armor, high speed, and small guns.
   (d) heavy armor, high speed, and medium guns.
   (e) light armor, high speed, and medium guns.

3. Cruisers are able to maintain high speed for a long time because
   (a) they have a wide beam.
   (b) they are divided into many small air-filled compartments.
   (c) they have only three gun turrets.
   (d) they carry a large supply of fuel oil.
   (e) they do not have armor plate.

4. In order to make it more difficult to sink cruisers, they are
   (a) equipped with thick armor.
   (b) armed with large guns.
   (c) fitted with mine-laying equipment.
   (d) powered by turbines with reduction gear.
   (e) divided into sections that can be flooded without affecting other parts.

DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]

After the setting-up exercises, the men are dismissed and then they must get into uniform for the drill which is to follow. Each man should have a uniform handy, either in his bag or locker, so that he can change into the proper uniform for drill in a few minutes. When drill call is sounded, men go quickly to their stations for drill and keep silence. A slow and noisy division is always inefficient. Drills are held to train all hands into an efficient fighting unit. All drills are for this one purpose. If you fail to act in drill as you should in an emergency, you will not know what to do when an emergency comes.

5. When you hear the drill call, you should
   (a) report to your officer.
   (b) put on your drill uniform.
   (c) remain where you are and keep silence.
   (d) find out where the drill is to be.
   (e) go at once to your station.

6. Which one of the following is not mentioned in the paragraph?
   (a) fighting.
   (b) punishment.
   (c) exercises.
   (d) drills.
   (e) uniforms.

7. If a division is slow and noisy
   (a) it will be dismissed.
   (b) it is not neatly dressed.
   (c) it will behave better in an emergency.
   (d) it cannot drill well.
   (e) the men will be punished.

8. The only thing the paragraph tells you about the clothing you wear at setting-up exercises is that
   (a) it is not drill uniform.
   (b) it should always be clean and neat.
   (c) you should be able to find it quickly.
   (d) it allows you to move freely.
   (e) it is kept in a locker.

9. The most important thing about your drills is that they
   (a) make you more alert.
   (b) make you strong.
   (c) train you for battle.
   (d) are required for all enlisted men.
   (e) teach you everything you should know.

DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]

If you neglect to obey an order or refuse to obey it, your chances of being sustained are less than one in a thousand. If an officer or petty officer orders you to do a certain thing, and before it is done, another officer or petty officer orders you to do something else, it is your duty to inform the officer or petty officer giving you the second order that you have had previous orders, telling him what the orders are and who gave them to you. The officer or petty officer who gave you the second order will then decide whether you are to carry out the first orders or the second orders.

10. The paragraph talks mainly about
    (a) the importance of good judgment.
    (b) how officers give orders.
    (c) respect for officers.
    (d) obedience to orders.
    (e) self-reliance.

11. There are specific instructions in the paragraph on
    (a) courteous behavior to officers.
    (b) what to do if you receive conflicting orders.
    (c) how to carry out an order.
    (d) the decisions you will be expected to make.
    (e) how to decide whether the first or second officer is right.

12. Should you ever tell an officer giving a second order that it conflicts with previous orders?
    (a) yes, if it does not interfere with the immediate following out of orders which have been given by another officer.
    (b) only if he has a lower rank than the officer giving the first order.
    (c) only if the officer asks you.
    (d) the paragraph does not answer this question.
    (e) yes, you should always do this when you receive conflicting orders.

13. You will be justified if you
    (a) occasionally do not obey an order.
    (b) disobey an order only once in a thousand times.
    (c) decide not to obey an order because you have received a second order before you carried out the first.
    (d) always obey the second of two orders.
    (e) neglect to obey an order because another officer tells you not to obey.

14. Your responsibility to the officer who gives the first of two successive orders
    (a) is maintained until his order is carried out.
    (b) includes making him a report of the conflicting order.
    (c) is not specifically mentioned in this paragraph.
    (d) is discharged at the moment when a second order is received.
    (e) cannot be transferred to the officer issuing the second order.

DO NOT STOP. GO ON TO THE NEXT PAGE.

The highest deck extending from bow to stern is called the main deck. A partial deck above the main deck at the bow only and not covering amidships is called the forecastle deck. A partial deck at the stern only and not covering amidships is called the poop deck. A complete deck below the main deck is called the second deck. Where there are two or more complete decks below the main deck, they are called second deck, third deck, fourth deck, etc. A partial deck above the lowest complete deck and below the main deck is called a half deck. Decks which for protective purposes are fitted with plating of extra strength and thickness are defined for technical purposes as protective and splinter in addition to their regular names. Where there is only one such deck, it is defined as protective, and where there are two, that having the thicker plating is defined as splinter, in addition to the regular names.

15. The second deck is
    (a) above the main deck.
    (b) below the main deck.
    (c) not defined in this paragraph.
    (d) not a complete deck.
    (e) so-called because it is protective.

16. A partial deck above the main deck covering all except the bow is
    (a) called the upper deck.
    (b) not discussed in this paragraph.
    (c) called the poop deck.
    (d) called the second deck.
    (e) always a protective deck.

17. A splinter deck is
    (a) not discussed in this paragraph.
    (b) not for protective purposes.
    (c) a deck with extra plating but less heavily armored than the protective deck.
    (d) the same as the second deck.
    (e) a half deck which is usually but not always heavily armored and which is above the main deck.

18. The main deck is
    (a) always armor-plated.
    (b) always a splinter deck.
    (c) below the second deck.
    (d) the largest partial protective deck.
    (e) always a complete deck.

19. A deck extending from bow to stern above the main deck
    (a) is usually but not always for protective purposes.
    (b) is the forecastle deck.
    (c) does not exist.
    (d) is armor-plated.
    (e) is called the protective second deck.

20. If there were two decks of extra strength and thickness and that with the thinner plating was a complete deck immediately below the main deck, it would be called the
    (a) protective poop deck.
    (b) forecastle deck.
    (c) splinter half deck.
    (d) splinter second deck.
    (e) protective lower deck.

DO NOT STOP. GO ON TO THE NEXT PAGE.

When a magnetized needle is mounted on a horizontal axis so that the angle between the needle and a horizontal plane can be measured, it is called a dipping needle. The amount of dip varies according to the place on the earth where the dip is measured. In other words, the needle acts as though the earth itself were a huge magnet with a north and a south pole. At the north magnetic pole the needle will point directly downward. On a line roughly halfway between the north and south magnetic poles, the needle will be horizontal; this line is called the magnetic equator. At all other points the needle will be neither horizontal nor vertical, and the amount of dip will depend on the distance from the magnetic poles. However, the north magnetic pole of the earth is not identical with the north geographic pole, but is about 20 degrees south of it, a degree being equal to about 70 miles. When traveling on the earth's surface many points are found where the inclination or dip of the needle is the same. A line drawn through these points on a map is called an isoclinic line. The isoclinic line drawn through points of zero dip, that is, where the needle is horizontal, is called an aclinic line.

21. The statement that a map of the world indicating isoclinic lines would show only one aclinic line
    (a) is false.
    (b) cannot be made on the basis of information given above.
    (c) can be inferred from information contained in the paragraph.
    (d) means that there is only one point on the earth's surface where a dipping needle would read zero.
    (e) implies that the aclinic line and the geographic equator are the same.

22. A dipping needle at the north geographic pole would
    (a) point in a horizontal direction.
    (b) point in a nearly vertical direction.
    (c) point straight downward.
    (d) show about the same dip as at the equator.
    (e) show a small angle with the horizontal.

23. Isoclinic lines on a map of the world would
    (a) have a general direction which cannot be even approximately determined from the above paragraph.
    (b) tend to run north and south.
    (c) tend to run east and west.
    (d) be very far apart.
    (e) cross one another in an unpredictable way.

24. The magnetic equator
    (a) is the same as the geographic equator.
    (b) passes through the north and south magnetic poles.
    (c) is an aclinic line.
    (d) is neither an aclinic nor an isoclinic line.
    (e) is not defined in the above paragraph.

25. It can be inferred from the above paragraph that the magnetic and geographic equators
    (a) will intersect at only one point.
    (b) are identical.
    (c) are parallel lines.
    (d) bear no fixed relationship to each other.
    (e) will intersect at two points.

 [redacted]

Electric current which is continually changing its magnitude and reversing its direction of flow in a conductor is called alternating current; and may be detected either by measuring the amount of current or by measuring the electro-motive-force (e.m.f.) exerted by it. When either the alternating current or the e.m.f. has passed from zero to its maximum value in one direction, back to zero, then to its maximum value in the other direction and back to zero, it is said to have completed a cycle. The time required for the current or e.m.f. to pass through one complete cycle is called a period. When the periods are all of the same length, the current produced is called periodic current. If the current goes through the complete cycle 60 times per second, it is said to have a frequency of 60 cycles. Each complete cycle may be thought of as containing or measuring 360 degrees. A half-cycle or the change from zero to maximum in either direction and back to zero would be 180 degrees. If the current reaches a maximum before the e.m.f., the current is said to be leading the e.m.f. in phase. If the current reaches a maximum after the e.m.f., the current is said to lag in phase. In either case, they are said to be out of phase.

26. A current is not called alternating unless
    (a) reversals of direction of flow are instantaneous.
    (b) its changes in size or amount are large.
    (c) its magnitude is always changing.
    (d) its frequency is 60 cycles per second.
    (e) it is in phase with its e.m.f.

27. Current which is continually reversing its direction of flow but which requires a different length of time to complete each cycle
    (a) is called periodic alternating current.
    (b) is not alternating current.
    (c) does not exist.
    (d) is alternating but not periodic current.
    (e) is called direct current.

28. If a periodic alternating current has a frequency of 100 cycles, the period for this current
    (a) will be 1/100 of a second.
    (b) cannot be determined without additional information.
    (c) cannot be inferred solely from information given in the paragraph.
    (d) will be one hundred seconds.
    (e) is sometimes, but not necessarily, constant.

29. If the current and e.m.f. in a conductor are in such a relationship that one reaches a maximum in one direction when the other is at a maximum in the opposite direction, then
    (a) the current and e.m.f. are said to be in phase.
    (b) it can be concluded that no current will flow in the conductor.
    (c) they can be said to be 90 degrees apart in phase.
    (d) the current is leading the e.m.f. by three-fourths of a cycle.
    (e) the current and e.m.f. are 180 degrees apart in phase.

30. The statement that one-half of a cycle is called an alternation
    (a) is not true.
    (b) is not to be derived from the above paragraph.
    (c) means that the change in one direction is always equal in amount to the change in the opposite direction.
    (d) could be inferred from statements made in the paragraph.
    (e) means that the number of cycles per second is equal to the frequency.

STOP. WAIT FOR FURTHER INSTRUCTIONS.

[redacted]

TEST II

A R I T H M E T I C A L   R E A S O N I N G

Directions and Samples

In this test you will be given some problems in arithmetic. Following each problem are five answers, lettered from (a) to (e); one of these answers is correct. Your task is to solve each problem and black in the space on the separate answer sheet which corresponds to the answer you think is correct. The following problems are samples.

D. A boy bought a 25¢ War Stamp every day for 4 days. How much did he invest in Stamps?

(a) $0.50 (b) $1.00 (c) $1.50 (d) $2.00 (e) $2.50

The correct answer to the problem is $1.00, which is answer (b). Notice in row D on the separate answer sheet in the section labeled "Arithmetical Reasoning -- Samples" that space (b) has already been blacked in.

Now work the next problem, and in row E on the separate answer sheet black in the space which corresponds to the correct answer.

E. A ball team played 27 games and won 16 of them. How many did it lose?

(a) 1 (b) 7 (c) 10 (d) 11 (e) 43

Solve the next problem, and in row F on the answer sheet black in the space which corresponds to the correct answer.

F. If an airplane makes a trip of 500 miles in 2 hours, what is its average rate in miles per hour?

(a) 250 (b) 251 (c) 498 (d) 502 (e) 1,000

Now <u>check</u> your answers as I give the correct ones to you.

ARE THERE ANY QUESTIONS?

When the signal is given you will work more problems of the same kind. Mark each answer in the proper space on the separate answer sheet. Mark only one answer to each problem. Do not make any marks in this booklet. Work as fast and as accurately as you can.

Work every problem. If you are not sure, mark the answer you think is correct.

STOP HERE. DO NOT GO TO THE NEXT PAGE
UNTIL THE SIGNAL "READY BEGIN" IS GIVEN.

ARITHMETICAL REASONING

Test Problems

Solve each problem. Then black in the space corresponding to the correct answer.

31. The U. S. Navy proposes to have 700 combatant ships by a given date. If the Navy has 300 of the ships now, how many more will be needed?

    (a) 300 (b) 400 (c) 200 (d) 500 (e) 1,000

32. During World War I, keels were laid at Hog Island at the rate of one every 5 days. At that rate, how many keels were laid in 275 days?

    (a) 51 (b) 55 (c) 75 (d) 270 (e) 1375

33. A speed of 6 miles per minute is equivalent to how many miles per hour?

    (a) 10 (b) 60 (c) 100 (d) 240 (e) 360

34. A $50 war bond costs $37.50. How many of these bonds may be bought for $150?

    (a) 1 (b) 2 (c) 3 (d) 4 (e) 5

35. X's bank account is overdrawn $20.50. How much must he deposit to have a balance of $100 in his favor in the bank?

    (a) $79.50 (b) $80.50 (c) $119.50 (d) $120.50 (e) $125.00

36. A train leaving New York at 11:15 A.M. is scheduled to make the trip to Philadelphia in 1 hour and 53 minutes. When is it due to arrive?

    (a) 12:08 P.M. (b) 12:38 P.M. (c) 12:41 P.M. (d) 1:03 P.M. (e) 1:08 P.M.

37. What will be the cost for gasoline on a trip of 180 miles if the car averages 15 miles per gallon and gasoline costs 20 cents a gallon?

    (a) $1.35 (b) $1.67 (c) $2.40 (d) $2.50 (e) $54.00

38. London and Cologne, 330 miles apart by air, will be how many inches apart on a map whose scale is 50 miles to one inch?

    (a) .15 (b) 3.1 (c) 5 (d) 6.6 (e) 16.50

39. A "plebe" makes $780 a year at Annapolis. If he receives $300 immediately, and the rest is paid in 12 equal monthly installments, how much is each monthly installment?

    (a) $40.00 (b) $45.50 (c) $57.60 (d) $65.00 (e) $90.00

[redacted]

40. A destroyer travelled for two of three hours at full speed and then reduced its speed by 15 nautical miles during the third hour. If the total distance covered was 81 nautical miles what was its full speed in nautical miles per hour?

(a) 22 (b) 27 (c) 30 (d) 32 (e) 35

41. A destroyer division consists of 4 ships. Two divisions plus one leader make a squadron. Four squadrons make a flotilla. How many ships are there in a flotilla?

(a) 20 (b) 32 (c) 33 (d) 36 (e) 40

42. A propeller, with an initial speed of 25 revolutions per second, decreases its speed by 4 revolutions per second each 10 seconds. How many revolutions per second will it be making at the end of 30 seconds?

(a) 0 (b) 1 (c) 4 (d) 13 (e) $23\frac{2}{3}$

43. In an armor-piercing shell 3% of the total weight of the shell is bursting charge. What is the weight, in pounds, of the bursting charge in a 2,100 pound shell?

(a) 7 (b) 63 (c) 300 (d) 630 (e) 700

44. A cruiser can fire a total of 165 rounds a minute from its 15 six-inch guns. If six of its guns are "knocked out" how many rounds per minute can be fired?

(a) 66 (b) 75 (c) 90 (d) 99 (e) 135

45. If 3 parts sand by weight are mixed with 4 parts cement, what fraction of the total dry mixture is sand?

(a) $\frac{3}{7}$ (b) $\frac{4}{7}$ (c) $\frac{1}{4}$ (d) $\frac{3}{4}$ (e) $\frac{4}{3}$

46. Carrier A has 80 planes and carrier B has 75 planes. If $\frac{3}{8}$ of the planes on A and $\frac{1}{3}$ of the planes on B are disabled, how many planes are left to protect the carriers?

(a) 45 (b) 55 (c) 100 (d) 110 (e) 135

47. If, on the average, 4% of shipped fruit is spoiled, how many boxes should be shipped if 48 boxes of good fruit are wanted?

(a) $48\frac{1}{2}$ (b) 49 (c) 50 (d) 52 (e) 60

48. How many yards of material 36 inches wide will it take to black out a window measuring 2 yards wide and 3 yards long? (Assume no overlap of material.)

(a) 2 (b) 3 (c) 5 (d) 6 (e) $6\frac{2}{3}$

DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]

49. In 1937, 6.2 cents of the railroad dollar went for fuel. What was the fuel cost per thousand dollars income?

(a) $.62 (b) $6.20 (c) $62.00 (d) $161.00 (e) $620.00

50. The number of sailors in a certain station increased from 5 men to 30,000 men in 5 weeks. What was the average increase per week in the number of sailors?

(a) 1,200 (b) 5,991 (c) 5,999 (d) 6,000 (e) 6,001

51. If figs contain 78% water by weight, how much will 50 pounds of fresh figs weigh when dried? (Assume all water is removed.)

(a) 10 lbs. (b) 11 lbs. (c) 13 lbs. (d) 39 lbs. (e) 40 lbs.

52. A gap of 1 5/8 inches is to be filled in with four layers of metal. What will be the thickness in inches of the fourth layer if the other three are respectively 7/8 inch, 1/3 inch, and .25 inch thick?

(a) 1/3 (b) 1/4 (c) 1/6 (d) 1/8 (e) .25

53. The capacity of a standard coach is 80. If a troop train of 10 standard coaches is 20% overloaded, how many more cars would be needed to accommodate all the troops without overloading?

(a) 1.6 (b) 2 (c) 2.5 (d) 4 (e) 6

54. The approximate weight in pounds of a shell is equal to one-half the cube of the diameter (measured in inches) of the gun. What must the diameter of a gun be to fire a shell weighing 32 pounds?

(a) $2\sqrt[3]{2}$ in. (b) 4 in. (c) $4\sqrt[3]{4}$ in. (d) 8 in. (e) $21\frac{1}{3}$ in.

55. Water pressure P (in pounds per square foot) varies directly with depth $\underline{h}$ (measured in feet), according to the formula $P = hDg$, where $g = .003$. If the density D of water is 64 pounds per cubic foot, what is the pressure in pounds per square foot on a diver 20 feet below the surface?

(a) .01 (b) .38 (c) 3.84 (d) 9.6 (e) 104.2

56. Fifty-four ships of 5,000 tons each are required to carry 2 British divisions. If a 25,000 ton ship holds 5 times as much as a 5,000 ton ship, how many 25,000 ton ships are needed to carry 15 British divisions?

(a) 36 (b) 81 (c) 324 (d) 405 (e) 648

57. A plane makes a trip of 1,800 miles in 6 hours. How many hours would it have taken if the speed had been increased by $\frac{1}{3}$?

(a) 4 (b) $4\frac{1}{2}$ (c) 8 (d) 3 (e) $3\frac{1}{2}$

DO NOT STOP. GO ON TO THE NEXT PAGE.

58. A patrol bomber flies due south from its base for 2 hours at 300 miles per hour and then due east for 4 hours at 200 miles per hour. At how many miles per hour must he fly in going directly back to base to reach there in 5 hours?

(a) 200 (b) 250 (c) $250\sqrt{10}$ (d) 266 2/3 (e) 280

59. Countries X, Y, and Z were to be allotted all planes produced in November in the ratio 2:3:5. On that plan the quota for country X was 60 planes. If the allotment ratio for these planes were changed to 1:1:1, how many more planes would X receive?

(a) 60 (b) 20 (c) 30 (d) 40 (e) 90

60. Ship A leaves port at 12:00 noon and approaches position X, 240 miles away, at 12 miles per hour. Ship B, 1080 miles from X, leaves 2 hours later. At how many miles per hour must ship B travel in order to reach position X at the same time as ship A?

(a) $49\frac{1}{11}$ (b) 54 (c) 60 (d) 63 (e) 81

STOP. WAIT FOR FURTHER INSTRUCTIONS.

UNITED STATES NAVY

T E S T S   O F   R E A D I N G

A N D

A R I T H M E T I C A L   R E A S O N I N G

FORM 2

N A V P E R S-16512

1943

NOTICE:
This test is not to be shown, or the contents revealed, to unauthorized persons in or out of the Navy, or reproduced in whole or in part without written authorization from the Bureau of Naval Personnel of the United States Navy. This test booklet must be surrendered with the answer sheet at the close of the test.

DO NOT OPEN THIS BOOKLET UNTIL YOU ARE TOLD TO DO SO.

[redacted]

TEST I

R E A D I N G

Directions and Samples

In this test you will be given several paragraphs to read. Following each paragraph are a number of incomplete sentences, each of which is followed by five phrases lettered from (a) to (e). One of the five phrases added to the incomplete sentence makes a statement which according to the paragraph is true. After reading the paragraph, you are to decide which phrase makes each statement true. Then black in the space on the separate answer sheet which corresponds to that phrase.

Read the following sample paragraph; then answer question A.

> After a can of paint has been opened and the paint partly used, the can should be covered and kept as air-tight as possible to prevent a paint scum from forming on the surface. If scum forms, the paint should be strained through a fine-mesh wire screen or cheesecloth.

A. To prevent scum from forming in a partly-used can of paint, one should
 (a) keep the paint free from dirt.
 (b) fill the can up with water.
 (c) stir the paint well before storing.
 (d) keep the can tightly covered.
 (e) make sure that the can is more than half full.

The paragraph tells you that the can should be covered and kept as air-tight as possible to prevent scum from forming; so (d) is the phrase which completes the statement correctly. Therefore (d) has been blacked in on the separate answer sheet opposite A in the section labeled "Reading -- Samples".

Now answer question B and black in the space in line B on the answer sheet which corresponds to the phrase which completes the statement correctly, according to the paragraph.

B. The method of removing scum from paint is
 (a) to stir the paint thoroughly.
 (b) to pour the scum off.
 (c) to strain the paint through a wire screen.
 (d) to skim the scum off with a putty knife.
 (e) not discussed in the paragraph.

[redacted]

Answer question C and black in the space on the answer sheet which corresponds to the correct phrase.

C. The method of thinning paint is
   (a) to mix it with turpentine.
   (b) to mix it with linseed oil.
   (c) to mix it with white lead.
   (d) to mix it with water.
   (e) not discussed in the paragraph.

Now check your answers while I give the correct ones to you.

ARE THERE ANY QUESTIONS?

When the signal is given, you will have more paragraphs to read and questions to answer. Mark each answer in the proper space on the separate answer sheet. Mark only one answer to each question. Do not make any marks in this booklet. Work as fast and as accurately as you can.

Answer every question. If you are not sure of the answer, make the best guess you can.

DO NOT TURN THIS PAGE UNTIL THE SIGNAL "READY - BEGIN" IS GIVEN.

READING

Test Questions

Mark <u>only</u> on separate answer sheet.

Success in battle, the primary aim in every military organization, requires implicit obedience to orders. It requires that men be trained to do habitually everything that must be done in battle when under the fire of the enemy. Briefly, discipline is the habit of obedience by which a man obeys an order naturally and without question, without stopping to consider whether he wants to obey it or not; he must learn to obey simply because the order comes from higher authority.

1. Orders must be obeyed whenever
   (a) it is the sensible thing to do.
   (b) obedience protects the safety of others.
   (c) they are reasonable.
   (d) speed is essential.
   (e) they come from a person of higher rank.

2. The subject of this paragraph is
   (a) the speed with which a man must learn to do his job.
   (b) the way in which orders are given.
   (c) the necessity of doing immediately what you are told to do.
   (d) the consequences of disobedience.
   (e) the form of a military organization.

3. In obeying an order, it is important to
   (a) disregard your own immediate wishes.
   (b) consider carefully whether the order should be obeyed
   (c) do it deliberately and cautiously.
   (d) know that the order was proper.
   (e) consider whether it involves danger to others.

4. The final object of requiring instant obedience to orders is to
   (a) avoid argument.
   (b) maintain authority in the hands of the officers.
   (c) win battles.
   (d) enable men to get along with one another.
   (e) maintain military discipline.

DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]

An important task on a naval vessel is to supervise the firing of guns. This is the task of the chief fire-control officer (C.F.C.O.). The C.F.C.O. is in command of all subdivisions of the armament, but he normally performs only supervisory functions in connection with all but the main battery; the latter he directly controls. The fire-control tower is the C.F.C.O.'s battle station. This station is equipped with the necessary fire-control instruments for directing and observing the fire of the main battery. It is also equipped with the necessary communications so that the C.F.C.O. may at all times communicate directly with the captain, with the officers in immediate control of the secondary, antiaircraft, and torpedo batteries, and with the officers in charge of subdivisions of the main battery, such as plotting-room officer, turret officers, and spotters. The C.F.C.O. keeps the captain informed as to the range and practicability of opening fire. He gives the commands "commence firing" and "cease firing," when directed by the captain. Prior to opening fire with the main battery he designates the fire distribution to be employed, the targets to be fired at, and the gun directors to be used. He supervises the salvo signals of the main battery and determines the rate of fire.

5. This paragraph concerns mainly
   (a) the number of batteries that should be fired at a target.
   (b) the purpose of equipment in the fire-control tower.
   (c) the knowledge required by the C.F.C.O.
   (d) the function of the captain in time of battle.
   (e) the duties of the C.F.C.O.

6. The relationship between the C.F.C.O. and the captain is such that
   (a) it would be necessary for the captain to communicate with the C.F.C.O., but not the reverse.
   (b) two-way communication between C.F.C.O. and captain is absolutely necessary.
   (c) the captain gives the commands to commence and cease firing upon the advice of the C.F.C.O.
   (d) the captain determines the distribution of fire of the main battery.
   (e) the captain and C.F.C.O. have different duties so that one does not need to confer with the other.

7. The correct description of what the C.F.C.O. does is that he
   (a) is in direct control of all the armament.
   (b) commands the ship during battle.
   (c) directs the handling of the ship.
   (d) supervises the use of all of the ship's armament in battle.
   (e) operates the instruments used to observe the results of fire.

8. One of the main duties of the C.F.C.O. is to
   (a) name the targets to be fired at.
   (b) control the secondary battery fire.
   (c) operate the gun directors.
   (d) control the speed of the ship during battle.
   (e) plot the ship's course.

9. The fire-control tower is evidently a kind of central office for
   (a) the plotting of the ship's course.
   (b) the coordination of information regarding targets and available gun-power.
   (c) the direction of secondary battery fire only.
   (d) the spotting of the results of the fire.
   (e) officers in charge of the various batteries so that they may direct their batteries from it.

DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]

The one great danger in a small boat when running before a surf (that is, traveling in the same direction as the breaking waves) is that of broaching-to, which means being thrown broadside-on in the surf. A boat broaches-to when running before the sea because she opposes little resistance to it and hence is carried before the wave. Broaching-to is most likely to occur when a boat is headed for shore, because the incoming waves or rollers, on reaching the boat, throw up the stern and as a consequence depress the bow. If she then has sufficient inertia (the greater the weight, the greater the inertia) to allow the wave to pass her, she will in succession pass through the descending, the horizontal, and the ascending positions as the crest of the wave passes successively her stern, her midships, and her bow. But if a boat on being caught by a heavy wave has not sufficient inertia to allow it to pass her, only the first of the three positions above described occurs. The stern is raised high in the air and the wave carries the boat before it on its front, sometimes with frightful velocity, and the boat may be thrown end-over-end or be turned broadside to the sea and capsized.

10. In running before a sea, the part of the boat which is lifted first as a wave comes along
    (a) is the stern.
    (b) is the bow.
    (c) is the middle section.
    (d) can not be determined from the paragraph.
    (e) would depend on the size of the wave.

11. According to the paragraph, a very light boat would
    (a) have great inertia because it would ride over the water and not be carried before it.
    (b) show little danger of capsizing when running through surf.
    (c) be difficult to handle in a calm sea.
    (d) have a large inertia because inertia is proportional to the square of the weight.
    (e) have very little inertia.

12. In a boat running before the sea, there is danger of a boat's broaching-to when
    (a) the boat is very heavy.
    (b) each wave passes the boat quickly by.
    (c) the waves are high but not breaking.
    (d) the wave pushes the boat ahead of it.
    (e) the sea is very deep.

13. One result of an increase in the weight of the boat is to
    (a) increase the danger of broaching-to.
    (b) cause the boat to be affected to a greater extent by the wind.
    (c) make the boat more likely to sink.
    (d) make the boat more responsive to the force of the waves.
    (e) make it less likely that the boat will broach-to.

14. The chief topic of the paragraph is
    (a) the effect of the weight of a boat on its maneuverability.
    (b) the method of preventing a boat from broaching-to.
    (c) the way to land a boat through heavy surf.
    (d) the action of heavy surf on a boat.
    (e) the power of heavy seas.

15. Since a light boat running before a broken sea opposes little resistance to it,
    (a) the boat can travel faster than the breaking waves.
    (b) the waves are likely to break over the bow of the boat.
    (c) the waves are likely to break over the stern of the boat.
    (d) the boat may be pushed along rapidly by the front of the wave.
    (e) the boat travels so much more slowly than the waves that it is likely to capsize.

Articles such as guns, furniture, and special tools which are part of the ship's equipage or which make the ship manageable, habitable, and serviceable as a naval vessel are called title-B articles. Lost or missing title-B articles are surveyed as soon as possible by an officer appointed by the commanding officer if the value of any one article does not exceed $100, or the total of identical articles does not exceed $100, and are surveyed by a board of three officers if values are greater. The survey officer or board makes a searching and exhaustive investigation of the circumstances and in every case determines the cause of the loss. When the loss is due to carelessness on the part of any person, the commanding officer will assign punishment as the case demands. Surveying officers or boards do not hold men responsible for loss due to casualties or stress of weather if witnesses can be found to prove the cause of loss.

16. One situation mentioned in the paragraph in which men would not be held responsible for loss is when
    (a) the loss is due to negligence.
    (b) the loss is due to theft.
    (c) several men are responsible and when witnesses cannot be found to prove responsibility.
    (d) the loss is due to heavy seas if witnesses testify to this.
    (e) the loss is due to unavoidable deterioration through use.

17. Lost or missing title-B articles not exceeding $100 in value are surveyed by
    (a) a Chief Petty Officer.
    (b) some officer with the rank of Lieutenant Commander or higher, who is appointed by the commanding officer.
    (c) a board of three officers.
    (d) an officer appointed by the commanding officer.
    (e) a board consisting of the commanding officer and two junior officers.

18. The paragraph states that a survey board investigates the loss of
    (a) all articles missing from the ship.
    (b) title-B articles which are valued at less than $100.
    (c) all title-B articles.
    (d) title-B articles valued at more than $100.
    (e) only those title-B articles lost due to negligence.

19. When the loss of title-B articles has been established as due to carelessness,
    (a) punishment is determined by the survey officer or board.
    (b) punishment usually consists of loss of pay.
    (c) punishment is assigned which is not described in the above paragraph.
    (d) articles are replaced by the man causing the loss.
    (e) articles are replaced only if valued at less than $100.

20. The reason for the loss of title-B articles is
    (a) established only when an individual is responsible.
    (b) always determined by a board of officers.
    (c) always established no matter what the circumstances.
    (d) rarely successfully established.
    (e) not established when due to stress of weather.

DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]

Even during the hottest periods of the year, the temperature rise of the water on the ocean surface is very small compared with the rise in temperature of the land under similar circumstances. There are several reasons why this is true. Except when the sun is nearly overhead, most of its radiation to the surface of the sea is reflected back into space. As the sun approaches the meridian, at which time it is directly overhead, less of its radiation is reflected and more of it is absorbed by the ocean water. However, the sun's radiation penetrates to a considerable depth before it is completely absorbed. Whereas on land only a few inches of the surface absorbs all the radiation and has its temperature considerably increased, on the sea several fathoms of water absorb the radiation and the temperature increase of the water is very small. Another factor which acts to prevent temperature rise of the ocean surface is that some of the heat received is used in evaporating water. The evaporation cools the water and increases the concentration of salt, both of which increase the density of the water. The denser water then sinks and is replaced by cooler water from below.

21. The reason for the fact that the sea water washing the coast is warmer in summer than in winter is
    (a) that the temperature change of the surface of the ocean is very small.
    (b) not discussed in the paragraph.
    (c) that the land warms the water.
    (d) that the ocean is very salty.
    (e) not known to those who have studied the way in which the temperature of the sea changes.

22. Evaporation of sea water
    (a) acts to raise the temperature of the water.
    (b) does not affect the rate of heating the water.
    (c) is so small as to be negligible.
    (d) occurs more rapidly at night than during the day.
    (e) tends to counteract the heating effect of the sun's rays.

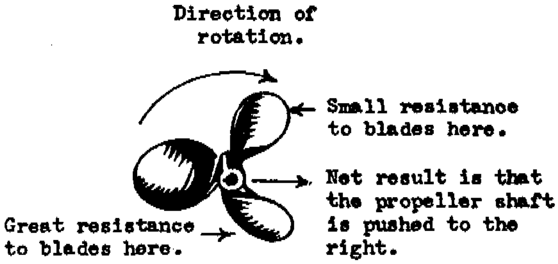
23. It can be concluded from statements made in the paragraph that
    (a) ocean temperatures are in general more stable than land temperatures.
    (b) land, being solid, retains heat longer and therefore its temperature is more stable than that of water.
    (c) because of the existence of currents in the ocean its temperature is less stable than land temperature.
    (d) the surface of the ocean heats more rapidly during the day than does the surface of the land.
    (e) the prediction of daily variations in ocean temperature is very difficult.

24. The formation of a layer of warm water tends to be prevented by
    (a) an increase in the density of surface water.
    (b) absorption of radiant energy which strikes the water at right angles.
    (c) absorption of heat from the air above the water.
    (d) the fact that water does not conduct heat well.
    (e) the fact that water is relatively incompressible.

25. One factor affecting the stability of land temperature as compared to that of sea water temperature is
    (a) the relative density of land and water.
    (b) seasonal variation in the concentration of salt in sea water.
    (c) the fact that the angle at which the rays strike the surface does not change on land as it does on water.
    (d) the fact that more solar radiation strikes the water than the land.
    (e) the fact that solar radiation penetrates water more deeply than it does land.

Viewed from a position behind the ship, a right-handed propeller turns clockwise when driving the ship ahead and a left-handed propeller turns counter-clockwise. Reversing the direction of rotation of the propeller drives the ship backward. When turning, the propeller exerts a sidewise force on the propeller shaft because the propeller blades pass through water under a greater pressure at the lower part of their arc than at the upper part of the arc. Thus there is a resultant force against the propeller shaft (and therefore the stern of the ship) opposite to the direction in which the lower blades travel.

Right-handed propeller viewed from rear.
Propeller turning clockwise.
Ship moving ahead.



26. In a ship with a single propeller, the relationship between the sidewise force exerted by a right-handed propeller and the force exerted by a rudder turned to the right
    (a) is such that the two forces would exactly counteract one another.
    (b) is such that the force exerted by the rudder would be greater than that exerted by the propeller.
    (c) cannot be determined from information given in the paragraph.
    (d) is such that the ship would move to the right.
    (e) would depend upon whether the ship is turning to the right or left.

27. Suppose a ship has two propellers, one on each side of the ship. If the sidewise force of one propeller is to counteract that of the other, then
    (a) the two propellers should always revolve in opposite directions.
    (b) both propellers should turn in the same direction.
    (c) it does not matter in what direction the propellers turn, provided they are the same size and turn at the same speed.
    (d) it does not matter in what direction the propellers turn, provided they turn at the same speed.
    (e) the direction in which the propellers should turn cannot be determined from the information given in the paragraph.

28. If a right-handed propeller is turning in such a direction as to cause the ship to move backward and no other force except that caused by the propeller is operating, the stern of the ship will
    (a) tend to be pushed to the right.
    (b) move directly backward.
    (c) move first to the right and then to the left.
    (d) move first to the left and then to the right.
    (e) tend to go to the left.

29. Assuming the same number of revolutions per minute, an increase in the diameter of the ship's propeller would
    (a) increase the sidewise force it would exert.
    (b) decrease the sidewise force it would exert.
    (c) have no effect on the tendency of the propeller to move the stern of the ship sidewise.
    (d) have an effect which cannot be determined from information given in the paragraph.
    (e) change the direction of the force exerted by the propeller in a way which would depend on the speed of the ship.

30. A right-handed propeller would, when turning clockwise, tend to
    (a) cause the ship to turn in a circle to the left.
    (b) cause the ship to move directly ahead.
    (c) cause the ship to turn in a circle to the right.
    (d) turn the ship in an unpredictable direction.
    (e) exert a directional force which cannot be determined from information given in the above paragraph.

STOP. WAIT FOR FURTHER INSTRUCTIONS.

[redacted]

TEST II

A R I T H M E T I C A L   R E A S O N I N G

Directions and Samples

In this test you will be given some problems in arithmetic. Following each problem are five answers, lettered from (a) to (e); one of these answers is correct. Your task is to solve each problem and black in the space on the separate answer sheet which corresponds to the answer you think is correct. The following problems are samples.

D. A boy bought a 25¢ War Stamp every day for 4 days. How much did he invest in Stamps.

(a) $0.50   (b) $1.00   (c) $1.50   (5) $2.00   (e) $2.50

The correct answer to the problem is $1.00, which is answer (b). Notice in row D on the separate answer sheet in the section labeled "Arithmetical Reasoning -- Samples" that space (b) has already been blacked in.

Now work the next problem, and in row E on the separate answer sheet black in the space which corresponds to the correct answer.

E. A ball team played 27 games and won 16 of them. How many did it lose?

(a) 1   (b) 7   (c) 10   (d) 11   (e) 43

Solve the next problem, and in row F on the answer sheet black in the space which corresponds to the correct answer.

F. If an airplane makes a trip of 500 miles in 2 hours, what is its average rate in miles per hour?

(a) 250   (b) 251   (c) 498   (d) 502   (e) 1,000

Now check your answers as I give the correct ones to you.

ARE THERE ANY QUESTIONS?

When the signal is given you will work more problems of the same kind. Mark each answer in the proper space on the separate answer sheet. Mark only one answer to each problem. Do not make any marks in this booklet. Work as fast and as accurately as you can.

Work every problem. If you are not sure, mark the answer you think is correct.

DO NOT TURN THIS PAGE UNTIL THE SIGNAL "READY - BEGIN" IS GIVEN.

[redacted]

ARITHMETICAL REASONING

Test Problems

Solve each problem. Then blacken the space corresponding to the correct answer.

31. One man out of every 6 from a company of 120 volunteered for scout duty. How many men volunteered?

    (a) 20 (b) 24 (c) 96 (d) 100 (e) 720

32. A train of 5 standard cars accommodates 400 passengers. If another standard car is added, how many passengers can be accommodated altogether?

    (a) 480 (b) 500 (c) 515 (d) 525 (e) 2,000

33. If a bomb used for practice purposes contains 5 pounds of black powder, 505 pounds of black powder will be sufficient for how many of these bombs?

    (a) 101 (b) 250 (c) 255 (d) 500 (e) 2,525

34. To strengthen each long ton of steel, 14 pounds of manganese are needed. How many pounds of manganese are needed to strengthen 26 long tons of steel?

    (a) $\frac{7}{13}$ (b) $1\frac{6}{7}$ (c) 12 (d) 40 (e) 364

35. There are 25 to 60 men in a platoon. A company consists of 3 to 4 platoons. What is the greatest number of men which can make up a company?

    (a) 75 (b) 149 (c) 170 (d) 240 (e) 595

36. If a food ration allows a person 7 ounces of bread a day, how many pounds is a person allowed per week? (1 pound = 16 ounces)

    (a) 1 (b) $3\frac{1}{16}$ (c) $3\frac{1}{2}$ (d) 16 (e) 784

37. A pace is $2\frac{1}{2}$ feet. How many paces are necessary to measure 150 feet?

    (a) 60 (b) 100 (c) 167 (d) 225 (e) 375

38. A canal, known to be 20 feet wide at point X, measures 1 inch wide in an aerial photograph. A military target lying near X measures 2 inches in diameter in the photograph. What is the actual diameter, in feet, of the target?

    (a) 10 (b) 30 (c) 40 (d) 80 (e) 40π

39. If $\frac{1}{5}$ of a ton of coal per day is required to heat a building, how many days will 20 tons last?

    (a) 4 (b) 10 (c) 25 (d) 75 (e) 100

[redacted]

40. In a certain electric company a worker's yearly pay has risen from $2,000 to $4,872 in the past 10 years. What is the average yearly increase?

(a) $14.36 (b) $28.72 (c) $287.20 (d) $687.20 (e) $1,436.00

41. If a mechanized division, averaging 30 miles per hour, leaves its base at 2:10 P.M., how many miles will it have gone by 5:30 P.M.?

(a) 9 (b) 39 (c) 96 (d) 100 (e) 200

42. Two towns, 15 miles apart, are $\frac{3}{4}$ inches apart on a map. On that map, one mile is represented by how many inches?

(a) $\frac{3}{100}$ (b) $\frac{1}{20}$ (c) $\frac{1}{2}$ (d) $11\frac{1}{4}$ (e) 20

43. The weight of the container is 6% of the shipping weight of a certain foodstuff. How many tons of the foodstuff can be shipped if the total shipping weight is to be 500 tons?

(a) 200 (b) 470 (c) 530 (d) 531.9 (e) 8,333

44. The rate of absenteeism has doubled since the start of the war. If 6 out of every 200 workers were absent before the war, what percentage of the workers are absent now?

(a) $1\frac{1}{2}$% (b) 6% (c) $16\frac{2}{3}$% (d) $33\frac{1}{3}$ % (e) 94%

45. The gasoline consumption of a car at 60 miles per hour is double the consumption at 30 miles per hour. A motorist, traveling at 30 miles per hour, gets 20 miles on a gallon of gasoline. If gasoline costs 20 cents a gallon, how much does he save on gasoline for a 100 mile trip by traveling at the slower speed?

(a) $.50 (b) $1.00 (c) $2.00 (d) $3.00 (e) $4.00

46. The United States force at the Solomons consisted of 2 heavy cruisers, 3 light cruisers, and 8 destroyers. What fraction of the force was destroyers?

(a) $\frac{3}{4}$ (b) $\frac{8}{5}$ (c) $\frac{1}{6}$ (d) $\frac{5}{8}$ (e) $\frac{8}{13}$

47. Bomber X leaves London for Berlin one-half hour later than bomber Y. How many miles per hour must X travel to reach Berlin at the same time as Y if Y travels at 200 miles per hour and makes the trip in 3 hours?

(a) $66\frac{2}{3}$ (b) $166\frac{2}{3}$ (c) $171\frac{3}{7}$ (d) 240 (e) 267

48. Pre-enrollment examinations at a certain school eliminate 40% of the candidates for this school. Out of 150 candidates in 1941, how many were accepted into the school?

(a) 60 (b) 90 (c) 110 (d) 144 (e) 146

DO NOT STOP. GO ON TO THE NEXT PAGE.

[redacted]

49. A plane descends from an altitude of 1,000 feet at the rate of 12 feet each 5 seconds. What will be its altitude, in feet, at the end of 60 seconds?

(a) 144 (b) 280 (c) 856 (d) 975 (e) 999

50. A soldier drives his jeep at the rate of 20 miles per hour for the first 6 miles and at the rate of 40 miles per hour for the next 36 miles. How many hours does it take the soldier to drive the 42 miles?

(a) $\frac{7}{10}$ (b) $1\frac{17}{130}$ (c) $1\frac{11}{87}$ (d) $1\frac{1}{5}$ (e) $1\frac{2}{5}$

51. A gunnery crew, in firing at a target, makes 6 hits and 9 misses. What percentage of the shots are hits?

(a) $2\frac{1}{2}\%$ (b) $11\frac{1}{9}\%$ (c) 40% (d) 54% (e) $66\frac{2}{3}\%$

52. A transport plane can carry $\frac{1}{12}$ as much tonnage as a cargo ship of a certain class. If a flight consists of 4 transport planes, how many flights would be needed to carry the full load of 3 cargo ships?

(a) $\frac{7}{12}$ (b) 1 (c) 4 (d) 9 (e) 16

53. One officer and 8 men make up the crew of a patrol boat. What fraction of the total crew in a division of 4 patrol boats is officers?

(a) $\frac{1}{32}$ (b) $\frac{1}{9}$ (c) $\frac{1}{8}$ (d) $\frac{4}{9}$ (e) $\frac{1}{2}$

54. The number of men at a station increased from 2,400 to 2,700. This increase was an increase of what per cent?

(a) 1.1% (b) 8% (c) 9% (d) 11.1% (e) 12.5%

55. Booklets numbered from 750 to 1000 inclusive were missing. How many booklets were missing?

(a) 247 (b) 248 (c) 249 (d) 250 (e) 251

56. Three competing factories turn out a certain product in the ratio 3:4:5. If the production of the first factory is doubled while that of the other two factories remains the same, the production of the third factory will be what fraction of the new total output of all three factories?

(a) $\frac{1}{3}$ (b) $\frac{5}{16}$ (c) $\frac{5}{12}$ (d) $\frac{1}{2}$ (e) $\frac{5}{7}$

57. The time in seconds, t, required by an object to fall d feet is represented by the formula: $d = 16t^2$. What is the altitude, in feet, of an airplane if a bomb dropped from it takes 20 seconds to land?

(a) $\frac{1}{2}\sqrt{5}$ (b) $32\sqrt{5}$ (c) 320 (d) 5,120 (e) 6,400

58. From each bonus payment must be deducted 1% for Social Security, 4% for state tax, and 5% for Victory tax. How much bonus must an employee be paid if he is to receive $10.00 after deductions have been made?

(a) $9.00 (b) $11.00 (c) $11.11 (d) $19.00 (e) $100.00

59. The sides, top, and bottom of a shipping crate are each $\frac{1}{2}$ inch thick. If each outer edge of the crate is 1 foot long, what is the inner capacity of the crate in cubic inches?

(a) 121 (b) 1,296 (c) 1,331 (d) $1{,}520\frac{7}{8}$ (e) 1,728

60. A patrol plane sights a motionless target 6 miles due south and radios a battleship which is 5 miles due east of the plane. For what range, in miles, should the guns of the battleship be set to fire on the target?

(a) $5\frac{1}{2}$ (b) $\sqrt{11}$ (c) $\sqrt{30}$ (d) $\sqrt{61}$ (e) 61

STOP. WAIT FOR FURTHER INSTRUCTIONS.

INDEX

The subject indexes of all STR volumes are combined in a master index printed in a separate volume. For access to the index volume consult the Army or Navy Agency listed on the reverse of the half-title page.

[redacted]

[redacted]

[redacted]

[redacted]

UNCLASSIFIED RESTRICTED

ATI 32213

TITLE: Summary Technical Report of the Applied Psychology Panel, NDRC - Human Factors in Military Efficiency

AUTHOR(S) : Wolfle, Dael; Lindsley, D. B.; Kappauf, W. E.; and others

ORIG. AGENCY : OSRD, NDRC, Applied Psychology Panel, Washington, D. C.

PUBLISHED BY : (None)

REVISION (None)

ORIG. AGENCY NO. (None)

PUBLISHING AGENCY NO. (None)

DATE	DOC. CLASS.	COUNTRY	LANGUAGE	PAGES	ILLUSTRATIONS
'46	Restr.	U. S.	English	306	photos, tables, diagrs, graphs

ABSTRACT:

5/15/62

A systematic account of the work performed under the direction of the Applied Psychology Panel is presented. Selection and classification of military personnel, as well as military training and the human factors involved in the design and operation of military equipment are described.

DISTRIBUTION: Copies of this report obtainable from CADO.

DIVISION: Military Operations (24)
SECTION: Personnel and Training (3)

SUBJECT HEADINGS: Military personnel - Aptitude tests (62011.8)

UNCLASSIFIED

CONFIDENTIAL

ATI SHEET NO.: R-24-3-17

Central Air Documents Office
Wright-Patterson Air Force Base, Dayton, Ohio

AIR TECHNICAL INDEX

RESTRICTED

UNCLASSIFIED

RESTR.

ATI-	13214
REVISION	(None)
ORIG. AGENCY NO.	(None)
PUBLISHING AGENCY NO.	(None)

TITLE: Summary

AUTHOR(S) : Wolfle, Dael
ORIG. AGENCY : OSRD, NDRC, Applied Psychology Panel, Washington, D. C.
PUBLISHED BY : (None)

DATE	DOC. CLASS.	COUNTRY	LANGUAGE	PAGES	ILLUSTRATIONS
'46	Restr.	U. S.	English	9	table

8/15/62

ABSTRACT:

A summary of the work of the Applied Psychology Panel on problems of military selection and classification is given. The chapter is divided into four sections: development of tests for use in selecting military specialists, the Navy classification program, methods employed in the construction of tests for the selection and classification of military personnel, and the Panel's recommendations for future development of military classification processes.

DISTRIBUTION: Copies of this report obtainable from CADO.

DIVISION: Military Operations (24)
SECTION: Personnel and Training (3)

SUBJECT HEADINGS: Military personnel - Aptitude tests (62011.8)

ATI SHEET NO.: R-24-3-13

Central Air Documents Office
Wright-Patterson Air Force Base, Dayton, Ohio

AIR TECHNICAL INDEX

UNCLASSIFIED

RESTRICTED

RESTRICTED

TITLE: The United States Navy Basic Classification Test Battery	ATI- 13215
AUTHOR(S) : Wolfle, Dael	REVISION (None)
ORIG. AGENCY : OSRD, NDRC, Applied Psychology Panel, Washington, D. C.	ORIG. AGENCY NO. (None)
PUBLISHED BY : (None)	PUBLISHING AGENCY NO. (None)

UNCLASSIFIED

DATE	DOC. CLASS.	COUNTRY	LANGUAGE	PAGES	ILLUSTRATIONS
1946	Restr.	U.S.	English	17	tables

ABSTRACT:

5/15/62

When expansion of the U. S. Navy began, it became apparent that the tests and procedures being used to classify and assign enlisted personnel were out of date and inefficient. A new method for testing and classifying was inagurated and found superior. Comparison between test scores and grades made in a variety of Service schools showed that the tests predicted well which men would and which would not do well in differnt Service schools.

DISTRIBUTION: Copies of this report obtainable from CADO.

DIVISION: Military Operations (24)
SECTION: Personnel and Training (3)

SUBJECT HEADINGS: Classification tests (23301.8)

ATI SHEET NO.:

UNCLASSIFIED

Central Air Documents Office
Wright-Patterson Air Force Base, Dayton, Ohio

AIR TECHNICAL INDEX

RESTRICTED

RESTRICTED

TITLE: The Selection of Officers	ATI- 13216
	REVISION (None)
AUTHOR(S) :Wolfle, Dael	ORIG. AGENCY NO. (None)
ORIG. AGENCY :OSRD, NDRC, Applied Psychology Panel, Washington, D. C.	PUBLISHING AGENCY NO. (None)
PUBLISHED BY :(Same)	

DATE	DOC. CLASS.	COUNTRY	LANGUAGE	PAGES	ILLUSTRATIONS
1946	Restr.	U.S.	English	9	tables, graph

ABSTRACT:

The U. S. Navy officer qualification test was developed for appraising the qualifications of men applying for naval commissions. Several studies dealt with the selection of officers for special types of duty. Tests developed for radar operators were used in selecting tactical radar officers. The NROTC selection test was analyzed and recommendations for improvement were made. The value of various factors in predicting the success of amphibious flotilla and group commanders was investigated.

DISTRIBUTION: Copies of this report obtainable from CADO.

DIVISION: Military Operations (24)	SUBJECT HEADINGS: ~~Military personnel~~ - Aptitude tests (62011.8)
SECTION: Personnel and Training (3)	
ATI SHEET NO.: R-24-3-17	

Central Air Documents Office
Wright-Patterson Air Force Base, Dayton, Ohio

AIR TECHNICAL INDEX

RESTRICTED

RESTRICTED

UNCLASSIFIED

TITLE: Eliminating the Emotionally Unfit (Officer.)

ATI- 13217

REVISION (None)

ORIG. AGENCY NO. (None)

PUBLISHING AGENCY NO. (None)

AUTHOR(S) : Wolfle, Dael
ORIG. AGENCY : OSRD, NDRC, Applied Psychology Panel, Washington, D. C.
PUBLISHED BY : (Same)

DATE	DOC. CLASS.	COUNTRY	LANGUAGE	PAGES	ILLUSTRATIONS
1946	Restr.	U.S.	English	17	tables, graph

5/15/62

ABSTRACT:

The personal inventory was developed in order to have available a device for making quick identification of emotionally unstable men who are likely to break down under the stress of hazardous duty. This method consists of paired statements from which the man being tested selects the one which best fits him. Each pair includes a symptomatic and a nonsymptomatic statement. The number of symptomatic statements checked constitutes a man's score. Men making high scores were much more likely to be judged unfit for service, when examined by a psychiatrist, than were men making low scores.

DISTRIBUTION: Copies of this report obtainable from CADO.

DIVISION: Military Operations (24)
SECTION: Personnel & Training (3)

SUBJECT HEADINGS: Military personnel - Aptitude tests (62011.8)

ATI SHEET NO.: R-24-3-17

UNCLASSIFIED

Central Air Documents Office
Wright-Patterson Air Force Base, Dayton, Ohio

AIR TECHNICAL INDEX

RESTRICTED

UNCLASSIFIED

~~RESTRICTED~~

TITLE: Determining Vocational Interests	ATI- 13218
AUTHOR(S): Wolfle, Dael	REVISION (None)
ORIG. AGENCY: OSRD, NDRC, Applied Psychology Panel, Washington, D. C.	ORIG. AGENCY NO. (None)
PUBLISHED BY: (None)	PUBLISHING AGENCY NO. (None)

DATE	DOC. CLASS.	COUNTRY	LANGUAGE	PAGES	ILLUSTRATIONS
1946	Restr.	U.S.	English	5	table

ABSTRACT:

5/15/62

To secure information on a wide variety of interests such as vocational, social, recreational, scientific and religious, and inventory of interest was constructed for possible use in the classification of military personnel. Information of this type could provide a valuable supplement to that obtained by interviews, vocational history, or standardized tests of ability in classifying men for military duty.

DISTRIBUTION: Copies of this report obtainable from CADO.

DIVISION: Military Operations (24) SECTION: Personnel & Training (3)	SUBJECT HEADINGS: Military personnel - Aptitude tests (62011.8)
ATI SHEET NO.: R-24-3-17	UNCLASSIFIED

Central Air Documents Office
Wright-Patterson Air Force Base, Dayton, Ohio

AIR TECHNICAL INDEX

~~RESTRICTED~~

RESTRICTED UNCLASSIFIED CONFIDENTIAL U

TITLE: The Selection of Radio Code Operators

ATI- 13219

REVISION (None)

ORIG. AGENCY NO. (None)

PUBLISHING AGENCY NO. (None)

AUTHOR(S) :Wolfe, Dael

ORIG. AGENCY : OSRD, NDRC, Applied Psychology Panel, Washington, D.C.

PUBLISHED BY : (Same)

DATE	DOC. CLASS.	COUNTRY	LANGUAGE	PAGES	ILLUSTRATIONS
1946	Restr.	U.S.	English	5	tables, graphs

5/15/62

ABSTRACT:

At Navy request, the Applied Psochology Panel developed an improved test of radio code aptitude. The test first requires the subject to learn three simple characters of the radio code and then measures his ability to differentiate these characters as they are sent at faster and faster speeds. This new test was demonstrated to have a validity of .50 and to be superior to all other available code aptitude tests.

DISTRIBUTION: Copies of this report obtainable from CADO

DIVISION:Military Operations (24)

SECTION:Personnel and Training (3)

SUBJECT HEADINGS:

Military personnel - Aptitude tests (62011.8)

UNCLASSIFIED

ATI SHEET NO.:

Central Air Documents Office
Wright-Patterson Air Force Base, Dayton, Ohio

AIR TECHNICAL INDEX

RESTRICTED

UNCLASSIFIED

ATI- 13220	
DIVISION	(None)
ORIG. AGENCY NO.	(None)
PUBLISHING AGENCY NO.	(None)

TITLE: Selecting Radar Operators

AUTHOR(S) : Lindsley, D. B.
ORIG. AGENCY : OSRD, NDRC, Applied Psychology Panel, Washington, D.C.
PUBLISHED BY : (Same)

DATE	DOC. CLASS.	COUNTRY	LANGUAGE	PAGES	ILLUSTRATIONS
1946	Restr.	U.S.	English	6	tables

5/15/62

ABSTRACT:

A number of tests were designed for use in selecting and screening radar operators. The tests were intended to measure aptitude for the visual tasks of reading and interpreting oscilloscope patterns. Most of the tests showed satisfactory reliability. Validity data were unreliable and unsatisfactory because of the lack of an adequate criterion of proficiency for radar operators.

DISTRIBUTION: Copies of this report obtainable from CADO

DIVISION: Military Operations (24)
SECTION: Personnel & Training (3)

SUBJECT HEADINGS:
Military personnel - Aptitude tests (62011.8)

UNCLASSIFIED

ATI SHEET NO.:

Central Air Documents Office
Wright-Patterson Air Force Base, Dayton, Ohio

AIR TECHNICAL INDEX

RESTRICTED

RESTRICTED

UNCLASSIFIED

TITLE: Selective Stereoscopic Rangefinder and Heightfi[illegible]

AUTHOR(S) : Kappauf, W. E.
ORIG. AGENCY : OSRD, NDRC, Applied Psychology Panel, Washing[illegible]
PUBLISHED BY : (Same)

ATI- 13221

REVISION (None)

ORIG. AGENCY NO. (None)

PUBLISHING AGENCY NO. (None)

DATE	DOC. CLASS.	COUNTRY	LANGUAGE	PAGES	ILLUSTRATIONS
1946	Restr.	U.S.	English	24	photos, tables, diagrs, graphs

5/15/62

ABSTRACT:

Standards for the selection of stereoscopic heightfinder operators were developed, recommended to, and adopted by the Army. Assistance was given to the Army in setting up stereoscopic testing centers where men to be trained as heightfinder operators were selected. In the course of the work on improving selection standards, information was obtained on the reliability of a number of visual test instruments.

DISTRIBUTION: Copies of this report obtainable from CADO.

DIVISION: Military Operations (24)
SECTION: Personnel and Training (3)

SUBJECT HEADINGS:
Military personnel - Aptitude tests (62011.8)

UNCLASSIFIED

ATI SHEET NO.:

Central Air Documents Office
Wright-Patterson Air Force Base, Dayton, Ohio

AIR TEC[illegible]L INDEX

RES[illegible]CTED

RESTRICTED UNCLASSIFIED

ATI 06222 ATI 13222

REVISION (None)

ORIG. AGENCY NO. (None)

PUBLISHING AGENCY NO. (None)

TITLE: Selecting Night Lookouts

AUTHOR(S) : Kappauf, W. E.

ORIG. AGENCY : OSRD, NDRC, Applied Psychology Panel, Washington, D. C.

PUBLISHED BY : (Same)

DATE	DOC. CLASS.	COUNTRY	LANGUAGE	PAGES	ILLUSTRATIONS
1946	Restr.	U.S.	English	7	tables, diagrs

5/15/62

ABSTRACT:

The NDRC adaptometer, Model III, and the modified Rostenberg adaptometer were developed for night vision testing. Two studies on the prediction of night lookout performance both showed that the tests of night vision being used had low reliability and very low validity. Failure to achieve satisfactory methods for selecting night lookouts hinges on at least three factors - the unreliability of night vision tests, the unreliability of lookout performance measures, and the probable importance for lookout performance of traits other than those measured by the tests used.

DISTRIBUTION: Copies of this report obtainable from CADO.

DIVISION: Military Operations (24)
SECTION: Personnel and Training (3)

SUBJECT HEADINGS:

Military personnel - Aptitude tests (62011.8)

ATI SHEET NO.:

RESTRICTED

UNCLASSIFIED

TITLE: Selecting Navy Telephone Talkers

ATI- 13223

REVISION (None)

ORIG. AGENCY NO. (None)

PUBLISHING AGENCY NO. (None)

AUTHOR(S) : Wolfle, Dael
ORIG. AGENCY : OSRD, NDRC, Applied Psychology Panel, Washington, D.C.
PUBLISHED BY : (Same)

DATE	DOC. CLASS.	COUNTRY	LANGUAGE	PAGES	ILLUSTRATIONS
19 6	Restr.	U.S.	English	3	(None)

ABSTRACT:

5/15/62

To rate the potential ability of men as shipboard telephone talkers, a speech interview was developed. The interview was conducted over sound-powered telephones. It included the pronónciation of numbers and of all common American speech sounds, repetition of commands and extemporaneous description.

DISTRIBUTION: Copies of this report obtainable from CADO

DIVISION: Military Operations (24)
SECTION: Personnel & Training (3)

SUBJECT HEADINGS:
Military personnel - Aptitude Tests (62011.9)

ATI SHEET NO.:

Central Air Documents Office
Wright-Patterson Air Force Base, Dayton, Ohio

AIR TECHNICAL INDEX

UNCLASSIFIED

RESTRICTED

RESTRICTED

UNCLASSIFIED

TITLE: Principles and Devices for Military Classification

AUTHOR(S) : Kennedy, John L.
ORIG. AGENCY : OSRD, NDRC, Applied Psychology Panel, Washington, D.C.
PUBLISHED BY : (Same)

ATI- 13224

REVISION (None)

ORIG. AGENCY NO. (None)

PUBLISHING AGENCY NO. (None)

DATE	DOC. CLASS.	COUNTRY	LANGUAGE	PAGES	ILLUSTRATIONS
1946	Restr.	U.S.	English	15	photos, table, diagrs

ABSTRACT:

5/15/62

Mass classification procedures for naval personnel and the requirements for the success of a classification program are described. A detailed knowledge of the requirements of each military specialty is necessary. Tests used to determine a man's qualifications should be as independent of each other as possible; they must be validated. Procedures must be standardized, and the programs should be under the supervision of classification specialists. Devices to aid in securing reliable information for use in classification were developed and installed on several ships.

DISTRIBUTION: Copies of this report obtainable from CADO

DIVISION: ~~Military~~ Operations (24)
SECTION: Personnel and training (3)

SUBJECT HEADINGS:
Military Personnel - Aptitude tests (62011.8)

ATI SHEET NO.:

Central Air Documents Office
Wright-Patterson Air Force Base, Dayton, Ohio

AIR TECHNICAL INDEX

RESTRICTED

UNCLASSIFIED

RESTRICTED

UNCLASSIFIED

TITLE: Organization of a Classification Program for Recruits

AUTHOR(S) : Frederiksen, N.
ORIG. AGENCY : OSRD, NDRC, Applied Psychology Panel, Washington, D. C.
PUBLISHED BY : (Same)

ATI- 13225

REVISION (None)

ORIG. AGENCY NO. (None)

PUBLISHING AGENCY NO. (None)

DATE	DOC. CLASS.	COUNTRY	LANGUAGE	PAGES	ILLUSTRATIONS
1946	Restr.	U.S.	English	7	(None)

5/15/62

ABSTRACT:

The procedures followed in classifying and assigning enlisted personnel in the Navy are briefly described. Improved classification would result from using a primary test battery to separate the men of school quality from those should be assigned to general detail and a secondary test battery to determine the particular school to which each man in the first group should be assigned, and from improving the interview by standardizing the weighting assigned to various factors on which the interviewer's recommendation is based.

DISTRIBUTION: Copies of this report obtainable from CADO.

DIVISION: Military Operations (24)
SECTION: Personnel and Training (3)

SUBJECT HEADINGS:
Military personnel - Aptitude test s (62011.8)

ATI SHEET NO.:

Central Air Documents Office
Wright-Patterson Air Force Base, Dayton, Ohio

AIR TECHNICAL INDEX

UNCLASSIFIED

RESTRICTED

RESTRICTED

UNCLASSIFIED

TITLE: Organization of an Advanced Classificatic- Program

ATI- 13226

REVISION (None)

ORIG. AGENCY NO. (None)

PUBLISHING AGENCY NO. (None)

AUTHOR(S) : Frederiksen, N.
ORIG. AGENCY : OSRD, NDRC, Applied Psychology Panel, Washington, D. C.

PUBLISHED BY : (Same)

DATE	DOC. CLASS.	COUNTRY	LANGUAGE	PAGES	ILLUSTRATIONS
1946	Restr.	U.S.	English	9	(None)

ABSTRACT:

5/15/62

The methods employed by the Navy in advanced classification and reclassification of enlisted personnel are briefly described. The principal procedures and recommendations for the improvement of advanced classification methods are more extensive construction and use of objective methods of determining the actual proficiency with which a man can perform the duties of a particular billet; improvement in the methods of rating shipboard proficiency; and use of brief oral tests to determine how much a man knows about the details of a particular type of duty.

DISTRIBUTION: Copies of this report obtainable from CADO.

DIVISION: Military Operations (24)

SECTION: Personnel and Training (3)

SUBJECT HEADINGS:

Military personnel - Aptitude tests (62011.8)

UNCLASSIFIED

ATI SHEET NO.:

AIR TEC INDEX

RESTRICTED

TITLE:	Construction and Standardization of Group Tests
AUTHOR(S):	Frederiksen, N.
ORIG. AGENCY:	OSRD, NDRC, Applied Psychology Panel, Washington, D. C.
PUBLISHED BY:	(Same)

ATI- 13227

DIVISION (None)

ORIG. AGENCY NO. (None)

PUBLISHING AGENCY NO. (None)

UNCLASSIFIED

DATE	DOC. CLASS.	COUNTRY	LANGUAGE	PAGES	ILLUSTRATIONS
1946	Restr.	U.S.	English	13	table

5/15/62

ABSTRACT:

Procedures recommended for constructing aptitude tests for the selection of military personnel are outlined. The development of aptitude tests of the paper-and-pencil variety which are power tests rather than speed tests are described in detail. The best tests conducted under Applied Psychology Panel direction were those in which the recommended procedures were followed most closely. When short cuts were used, the tests suffered by comparison. Thoroughness in the application of test methods is strongly recommended.

UNCLASSIFIED

DISTRIBUTION: Copies of this report obtainable from CADO.

DIVISION: Education and Training (38)

SECTION: General (0)

SUBJECT HEADINGS:

Aptitude tests (11381)

Military personnel - Aptitude tests (62011.8)

ATI SHEET NO.:

Central Air Documents Office
Wright-Patterson Air Force Base, Dayton, Ohio

AIR TECH INDEX

RESTRICTED

RESTRICTED

TITLE: Contributions to Test Theory and Test Administration

AUTHOR(S) : Wolfle, Dael
ORIG. AGENCY : OSRD, NDRC, Applied Psychology Panel, Washington DC
PUBLISHED BY : (Same)

ATI- 13228
REVISION (None)
ORIG. AGENCY NO. (None)
PUBLISHING AGENCY NO. (None)

UNCLASSIFIED

DATE	DOC. CLASS.	COUNTRY	LANGUAGE	PAGES	ILLUSTRATIONS
1946	Restr.	U.S.	English	8	tables, diagr, graphs

ABSTRACT:

5/15/62

Methods of controlling and checking scoring procedures were devised because accuracy with which aptitude tests were being scored in 1942 revealed frequent and large errors. Subsequent investigations showed that errors were greatly reduced in frequency and size. Several inovations in test construction practice were tried out including a simple method of measuring the premium which a test places on speed.

UNCLASSIFIED

DISTRIBUTION: Copies of this report obtainable from CADO

DIVISION: Education and Training (38)
SECTION: General (0)

ATI SHEET NO.:

SUBJECT HEADINGS:
Aptitude tests (11381)
Military personnel - Aptitude tests (62011.8)

Central Air Documents Office
Wright-Patterson Air Force Base, Dayton, Ohio

AIR TECH INDEX

RESTRICTED